

□ **ROA Version, 8/2002.** Essentially identical to the Tech Report, with new pagination (but the same footnote and example numbering); correction of typos, oversights & outright errors; improved typography; and occasional small-scale clarificatory rewordings. Citation should include reference to this version.

# OPTIMALITY THEORY

## Constraint Interaction in Generative Grammar

First circulated: April, 1993  
RuCCS-TR-2; CU-CS-696-93: July, 1993  
Minor Corrections: December, 1993  
ROA Version: August, 2002

**Alan Prince**

Department of Linguistics  
Rutgers Cognitive Science Center  
Rutgers University  
prince@ruccs.rutgers.edu

**Paul Smolensky**

Department of Cognitive Science  
The Johns Hopkins University  
[1993: University of Colorado at Boulder]  
smolensky@cogsci.jhu.edu

Everything is possible but not  
everything is permitted ...

— Richard Howard, “The Victor Vanquished”

“It is demonstrated,” he said, “that things cannot be  
otherwise: for, since everything was made for a purpose,  
everything is necessarily made for the best purpose.”

— *Candide ou l’optimisme*. Ch. I.

*Remark.* The authors’ names are arranged in lexicographic order.

## Acknowledgments

Special thanks to John McCarthy for detailed discussion of virtually every issue raised here and for a fine-grained skepsis of the entire first draft of the ms., which resulted in innumerable improvements and would have resulted in innumerably more, were this a better world. We are particularly grateful for his comments and suggestions *in rē* Chs. 7 and 9. We also wish to thank Robert Kirchner, Armin Mester, and Junko Itô for remarks that have had significant impact on the development of this work, as well as David Perlmutter, Vieri Samek-Lodovici, Cheryl Zoll, Henrietta Hung, Mark Hewitt, Jane Grimshaw, Ad Neeleman, Diana Archangeli, Henry Churchyard, Doug Pulleyblank, Moira Yip, Tom Bever, Larry Hyman, Andy Black, Mike Jordan, Lauri Karttunen, René Kager, Paul Kiparsky, Mike Kenstowicz, Ellis Visch, András Kornai, Akin Akinlabi, Géraldine Legendre, Clayton Lewis, Merrill Garrett, Jim Martin, Clara Levelt, Mike Mozer, Maria Bittner, Alison Prince, Dave Rumelhart, Mark Liberman, Jacques Mehler, Steve Pinker, Daniel Büring, Katharina Hartmann, Joshua Legendre Smolensky, Ray Jackendoff, Bruce Hayes, Geoff Pullum, Gyanam Mahajan, Harry van der Hulst, William Labov, Brian McHugh, Gene Buckley, Will Leben, Jaye Padgett and Loren Billings. None of these individuals can be sensibly charged with responsibility for any errors that may have crept into this work.

To Merrill Garrett (Cognitive Science, University of Arizona, Tucson) and to the organizers of the Arizona Phonology Conference we are grateful for providing in April 1991 the first public forums for the presentation of the theory, which proved a significant stimulus to the cohering thereof. We would also like to thank audiences at our 1991 LSA Summer Institute course and at the Feature Workshop there, at WCCFL 1992, at the OTS (Utrecht), University of California at Berkeley (Phonology Laboratory), the University of Colorado at Boulder and the Boulder Connectionist Research Group, Rutgers University (New Brunswick and Piscataway), Brandeis University, the University of Pennsylvania (the Linguistics Department and the Institute for Research in Cognitive Science), Princeton University Cognitive Science Center, Stanford University (Phonology Workshop and Parallel Distributed Processing Seminar), the University of Rochester Cognitive Science Program, and the International Computer Science Institute of Berkeley CA.

Financial support was provided by a University of Colorado Faculty Fellowship, by research funds from Rutgers University and from the Rutgers Center for Cognitive Science, and, most crucially, by NSF SGER BNS-90 16806 without which the rigors of long-distance collaboration would have proved daunting indeed.

We remember Robert Jeffers with special appreciation for constructing the Rutgers environment that so greatly facilitated the progress of this work.

# Table of Contents

1. Preliminaries .....	1
1.1 Background and Overview .....	1
1.2 Optimality .....	4
1.3 Overall Structure of the Argument .....	7
<b>Part I Optimality and Constraint Interaction</b>	
Overview of Part I .....	10
2. Optimality in Grammar: Core Syllabification in Imdlawn Tashlhiyt Berber .....	11
2.1 The Heart of Dell & Elmedlaoui .....	11
2.2 Optimality Theory .....	17
2.3 Summary of discussion to date .....	22
3. Generalization-Forms in Domination Hierarchies I	
Blocking and Triggering: Profuseness and Economy .....	23
3.1 Epenthetic Structure .....	24
3.2 Do Something Only When: The Failure of Bottom-up Constructionism .....	28
4. Generalization-Forms in Domination Hierarchies II	
Do Something Except When: Blocking, or The Theory of Profuseness .....	33
4.1 Edge-Oriented Infixation .....	33
4.2 Interaction of Weight Effects with Extrametricality .....	38
4.2.1 Background: Prominence-Driven Stress Systems .....	38
4.2.2 The Interaction of Weight and Extrametricality: Kelkar's Hindi .....	41
4.3 Nonfinality and Nonexhaustiveness .....	44
4.3.1 Nonfinality and the Laws of Foot Form: Raw Minimality .....	49
4.3.2 Nonfinality and the Laws of Foot Form: Extended Minimality Effects .....	54
4.4 Summary of Discussion of the <i>Except When</i> Effect .....	59
4.5 Except meets Only: Triggering and Blocking in a Single Grammar .....	59
5. The Construction of Grammar in Optimality Theory .....	73
5.1 Construction of Harmonic Orderings from Phonetic and Structural Scales .....	73
5.2 The Theory of Constraint Interaction .....	74
5.2.1 Comparison of Entire Candidates by a Single Constraint .....	74
5.2.1.1 ONS: Binary constraints .....	75
5.2.1.2 HNUC: Non-binary constraints .....	78
5.2.2 Comparison of Entire Candidates by an Entire Constraint Hierarchy .....	79
5.2.3 Discussion .....	83
5.2.3.1 Non-locality of interaction .....	83
5.2.3.2 Strictness of domination .....	85
5.2.3.3 Serial vs. Parallel Harmony Evaluation and Gen .....	86
5.2.3.4 Binary vs. Non-binary constraints .....	88
5.3 Pāṇini's Theorem on Constraint Ranking .....	88

## Part II Syllable Theory

Overview of Part II .....	92
6. Syllable Structure Typology I: the CV Theory .....	93
6.1 The Jakobson Typology .....	93
6.2 The Faithfulness Interactions .....	95
6.2.1 Groundwork .....	95
6.2.2 Basic CV Syllable Theory .....	98
6.2.2.1 Onsets .....	99
6.2.2.2 Codas .....	102
6.2.3 The Theory of Epenthesis Sites .....	104
7. Constraint Interaction in Lardil Phonology .....	107
7.1 The Constraints .....	107
7.2 The Ranking .....	117
7.2.1 Some Ranking Logic .....	117
7.2.2 Ranking the Constraints .....	120
7.3 Verification of Forms .....	127
7.3.1 Consonant-Final Stems .....	128
7.3.2 Vowel Final Stems .....	132
7.4 Discussion .....	135
8. Universal Syllable Theory II:	
Ordinal Construction of C/V	
and Onset/Coda Licensing Asymmetry .....	139
8.1 Associational Harmony .....	144
8.1.1 Deconstructing HNUC: Berber, Take 1 .....	144
8.1.2 Restricting to Binary Marks .....	147
8.2 Reconstructing the C and V Classes:	
Emergent Parameter Setting <i>via</i> Constraint Ranking .....	152
8.2.1 Harmonic Completeness of Possible Onsets and Peaks .....	152
8.2.2 Peak- and Margin-Affinity .....	154
8.2.3 Interactions with PARSE .....	156
8.2.4 Restricting Deletion and Epenthesis .....	157
8.2.5 Further Necessary Conditions on Possible Onsets and Nuclei .....	158
8.2.6 Sufficient Conditions on Possible Onsets and Nuclei .....	160
8.3 The Typology of Onset, Nucleus, and Coda Inventories .....	165
8.3.1 The Typology of Onset and Nucleus Inventories .....	165
8.3.2 Onset/Coda Licensing Asymmetries .....	171
8.3.3 An Example: Berber, Take 2 .....	178
8.4 Simplifying the Theory by Encapsulating Constraint Packages .....	183
8.4.1 Encapsulating the Association Hierarchies .....	183
8.4.2 An Example: Berber, Take 3 .....	185
8.4.3 Sufficiency and Richness of the Encapsulated Theory .....	185

## Part III Issues and Answers in Optimality Theory

9. Inventory Theory and the Lexicon . . . . .	191
9.1 Language-Particular Inventories . . . . .	191
9.1.1 Harmonic Bounding and Nucleus, Syllable, and Word Inventories . . . . .	193
9.1.2 Segmental Inventories . . . . .	195
9.2 Universal Inventories . . . . .	202
9.2.1 Segmental Inventories . . . . .	202
9.2.2 Syllabic Inventories . . . . .	208
9.3 Optimality in the Lexicon . . . . .	209
10. Foundational Issues and Theory-Comparisons . . . . .	215
10.1 Thinking about Optimality . . . . .	215
10.1.1 Fear of Optimization . . . . .	215
10.1.2 The Reassurance . . . . .	215
10.2 The Connectionism Connection, and other Computation-based Comparisons . . . . .	217
10.2.1 Why Optimality Theory has nothing to do with connectionism . . . . .	217
10.2.2 Why Optimality Theory is deeply connected to connectionism . . . . .	218
10.2.3 Harmony Maximization and Symbolic Cognition . . . . .	219
10.3 Analysis of ‘Phonotactics+Repair’ Theories . . . . .	221
10.3.1 CV Syllable Structure and Repair . . . . .	224
10.3.2 General Structure of the Comparisons: Repair Analysis . . . . .	226
10.3.3 Persistent Rule Theory . . . . .	228
10.3.3.1 English Closed Syllable Shortening . . . . .	229
10.3.3.2 Shona Tone Spreading . . . . .	231
10.3.3.3 Summary . . . . .	233
10.3.4 The Theory of Constraints and Repair Strategies . . . . .	233
Appendix <i>&lt;incomplete&gt;</i> . . . . .	241
A.1 The Cancellation and Cancellation/Domination Lemmas . . . . .	241
A.2 CV Syllable Structure . . . . .	241
A.3 Pāṇini's Theorem on Constraint-ranking . . . . .	241
References . . . . .	243



# 1. Preliminaries

## 1.1 Background and Overview

As originally conceived, the *RULE* of grammar was to be built from a Structural Description delimiting a class of inputs and a Structural Change specifying the operations that altered the input (e.g. Chomsky 1962). The central thrust of linguistic investigation would therefore be to explicate the system of predicates used to analyze inputs — the possible Structural Descriptions of rules — and to define the operations available for transforming inputs — the possible Structural Changes of rules. This conception has been jolted repeatedly by the discovery that the significant regularities were to be found not in input configurations, nor in the formal details of structure-deforming operations, but rather in the character of the *output* structures, which ought by rights to be nothing more than epiphenomenal. We can trace a path by which “conditions” on well-formedness start out as peripheral annotations guiding the interpretation of rewrite rules, and, metamorphosing by stages into constraints on output structure, end up as the central object of linguistic study.

As the theory of representations in syntax has ramified, the theory of operations has dwindled in content, even to triviality and, for some, nonexistence. The parallel development in phonology and morphology has been underway for a number of years, but the outcome is perhaps less clear — both in the sense that one view has failed to predominate, and in the sense that much work is itself imperfectly articulate on crucial points. What is clear is that any serious theory of phonology must rely heavily on well-formedness constraints; where by ‘serious’ we mean ‘committed to Universal Grammar’. What remains in dispute, or in subformal obscurity, is the character of the interaction among the posited well-formedness constraints, as well as the relation between such constraints and whatever derivational rules they are meant to influence. Given the pervasiveness of this unclarity, and the extent to which it impedes understanding even the most basic functioning of the grammar, it is not excessively dramatic to speak of the issues surrounding the role of well-formedness constraints as involving a kind of conceptual crisis at the center of phonological thought.

Our goal is to develop and explore a theory of the way that representational well-formedness determines the assignment of grammatical structure. We aim therefore to ratify and to extend the results of the large body of contemporary research on the role of constraints in phonological grammar. This body of work is so large and various as to defy concise citation, but we would like to point to such important pieces as Kisseberth 1972, Haiman 1972, Pyle 1972, Hale 1973, Sommerstein 1974, where the basic issues are recognized and addressed; to Wheeler 1981, 1988, Bach and Wheeler 1981, Broselow 1982, Dressler 1985, Singh 1987, Paradis 1988ab, Paradis & Prunet 1991, Hulst 1984, Kaye & Lowenstamm 1984, Kaye, Lowenstamm & Vergnaud 1985, Calabrese 1988, Myers 1991, Goldsmith 1990, 1991, Bird 1990, Coleman 1991, Scobbie 1991, which all represent important strands in recent work; as well as to Vennemann 1972, Bybee 1972, 1985, Liberman 1975, Goldsmith 1976, Liberman & Prince 1977, McCarthy 1979, McCarthy & Prince 1986, Selkirk 1980ab, 1981, Kiparsky 1980, 1982, Kaye & Lowenstamm 1981, McCarthy 1981, 1986, Lapointe & Feinstein 1982, Cairns & Feinstein 1982, Steriade 1982, Prince 1983, 1990, Kager & Visch 1983, Hayes 1984, Hyman 1985, Dressler 1985, Wurzel 1985, Borowsky 1986ab, Itô 1986, 1989, Mester 1986, 1992, Halle & Vergnaud 1987, Lakoff 1988, in press, Yip 1988, Cairns 1988, Kager 1989, Visch 1989, Clements 1990, Legendre, Miyata, & Smolensky 1990ab, Mohanan 1991, in press, Archangeli & Pulleyblank 1992, Burzio 1992ab, Itô, Kitagawa & Mester 1992, Itô & Mester 1992 — a sample of work which offers an array of perspectives on the kinds of problems

we will be concerned with — some close to, others more distant from our own, and some contributory of fundamental representational notions that will put in appearances throughout this work (for which, see the local references in the text below). Illuminating discussion of fundamental issues and an interesting conception of the historical development is found in Goldsmith 1990; Scobbie 1992 reviews some recent work. The work of Stampe 1973/79, though framed in a very different way, shares central abstract commitments with our own; perhaps more distantly related are Chapter 9 of Chomsky & Halle 1968 and Kean 1975. The work of Wertheimer 1923, Lerdahl & Jackendoff 1983 (chs. 3 and 12), Jackendoff 1983 (chs. 7 and 8), 1987, 1991, though not concerned with phonology at all, provides significant conceptual antecedents; similarly, the proposals of Chomsky 1986, and especially 1989, 1992, though very different in implementation, have fundamental similarities with our own. Rizzi 1990, Bittner 1993, and Legendre, Raymond, & Smolensky 1993, and Grimshaw in prep., are among recent works in syntax and semantics that resonate with our particular concerns.

The basic idea we will explore is that Universal Grammar consists largely of a set of constraints on representational well-formedness, out of which individual grammars are constructed. The representational system we employ, using ideas introduced into generative phonology in the 1970's and 1980's, will be rich enough to support two fundamental classes of constraints: those that assess output configurations *per se* and those responsible for maintaining the faithful preservation of underlying structures in the output. Departing from the usual view, we do not assume that the constraints in a grammar are mutually consistent, each true of the observable surface or of some level of representation. On the contrary: we assert that the constraints operating in a particular language are highly conflicting and make sharply contrary claims about the well-formedness of most representations. The grammar consists of the constraints together with a general means of resolving their conflicts. We argue further that this conception is an essential prerequisite for a substantive theory of UG.

It follows that many of the conditions which define a particular grammar are, of necessity, frequently violated in the actual forms of the language. The licit analyses are those which satisfy the conflicting constraint set *as well as possible*; they constitute the optimal analyses of underlying forms. This, then, is a theory of optimality with respect to a grammatical system rather than of well-formedness with respect to isolated individual constraints.

The heart of the proposal is a means for precisely determining which analysis of an input *best satisfies* (or least violates) a set of conflicting conditions. For most inputs, it will be the case that every possible analysis violates many constraints. The grammar rates all these analyses according to how well they satisfy the whole constraint set and produces the analysis at the top of this list as the output. This is the *optimal* analysis of the given input, and the one assigned to that input by the grammar. The grammatically well-formed structures are those that are optimal in this sense.

How does a grammar determine which analysis of a given input best satisfies a set of inconsistent well-formedness conditions? Optimality Theory relies on a conceptually simple but surprisingly rich notion of constraint interaction whereby the satisfaction of one constraint can be designated to take absolute priority over the satisfaction of another. The means that a grammar uses to resolve conflicts is to rank constraints in a *strict dominance hierarchy*. Each constraint has absolute priority over all the constraints lower in the hierarchy.



Such prioritizing is in fact found with surprising frequency in the literature, typically as a subsidiary remark in the presentation of complex constraints.<sup>1</sup> We will show that once the notion of constraint-precedence is brought in from the periphery and foregrounded, it reveals itself to be of remarkably wide generality, the formal engine driving many grammatical interactions. It will follow that much that has been attributed to narrowly specific constructional rules or to highly particularized conditions is actually the responsibility of very general well-formedness constraints. In addition, a diversity of effects, previously understood in terms of the triggering or blocking of rules by constraints (or merely by special conditions), will be seen to emerge from constraint interaction.

Although we do not draw on the formal tools of connectionism in constructing Optimality Theory, we will establish a high-level conceptual rapport between the mode of functioning of grammars and that of certain kinds of connectionist networks: what Smolensky (1983, 1986) has called ‘Harmony maximization’, the passage to an output state with the maximal attainable consistency between constraints bearing on a given input, where the level of consistency is determined exactly by a measure derived from statistical physics. The degree to which a possible analysis of an input satisfies a set of conflicting well-formedness constraints will be referred to as the *Harmony* of that analysis. We thereby respect the absoluteness of the term ‘well-formed’, avoiding terminological confusion and at the same time emphasizing the abstract relation between Optimality Theory and Harmony-theoretic network analysis. In these terms, a grammar is precisely a means of determining which of a pair of structural descriptions is more *harmonic*. Via pair-wise comparison of alternative analyses, the grammar imposes a harmonic order on the entire set of possible analyses of a given underlying form. The actual output is the most harmonic analysis of all, the optimal one. A structural description is well-formed if and only if the grammar determines it to be the optimal analysis of the corresponding underlying form.

With an improved understanding of constraint interaction, a far more ambitious goal becomes accessible: to build individual phonologies directly from universal principles of well-formedness. (This is clearly impossible if we imagine that constraints must be surface- or at least level-true.) The goal is to attain a significant increase in the predictiveness and explanatory force of grammatical theory. The conception we pursue can be stated, in its purest form, as follows: Universal Grammar provides a set of highly general constraints. These often conflicting constraints are all operative in individual languages. Languages differ primarily in how they resolve the conflicts: in the way they rank these universal constraints in strict domination hierarchies that determine the circumstances under which constraints are violated. A language-particular grammar *is* a means of resolving the conflicts among universal constraints.

On this view, Universal Grammar provides not only the formal mechanisms for constructing particular grammars, it also provides the very substance that grammars are built from. Although we shall be entirely concerned in this work with phonology and morphology, we note the implications for syntax and semantics.

---

<sup>1</sup> One work that uses ranking as a systematic part of the analysis is Cole 1992; thanks to Robert Kirchner for bringing this to our attention.

## 1.2 Optimality

The standard phonological rule aims to encode grammatical generalizations in this format:

$$(1) \quad A \rightarrow B / C \text{---} D$$

The rule scans potential inputs for structures CAD and performs the change on them that is explicitly spelled out in the rule: the unit denoted by A takes on property B. For this format to be worth pursuing, there must be an interesting theory which defines the class of possible predicates CAD (Structural Descriptions) and another theory which defines the class of possible operations  $A \rightarrow B$  (Structural Changes). If these theories are loose and uninformative, as indeed they have proved to be in reality, we must entertain one of two conclusions:

(i) phonology itself simply doesn't have much content, is mostly 'periphery' rather than 'core', is just a technique for data-compression, with aspirations to depth subverted by the inevitable idiosyncrasies of history and lexicon; or

(ii) the locus of explanatory action is elsewhere.

We suspect the latter.

The explanatory burden can of course be distributed quite differently than in the re-write rule theory. Suppose that the input-output relation is governed by conditions on the well-formedness of the *output*, 'markedness constraints', and by conditions asking for the *exact preservation of the input* in the output along various dimensions, 'faithfulness constraints'. In this case, the inputs falling under the influence of a constraint need share no input-specifiable structure (CAD), nor need there be a single determinate transformation ( $A \rightarrow B$ ) that affects them. Rather, we generate (or admit) a set of candidate outputs, perhaps by very general conditions indeed, and then we assess the candidates, seeking the one that best satisfies the relevant constraints. Many possibilities are open to contemplation, but some well-defined measure of value excludes all but the best.<sup>2</sup> The process can be schematically represented like this:

### (2) Structure of Optimality-theoretic grammar

$$\begin{array}{ll} \text{a. Gen}(\text{In}_k) & \rightarrow \{ \text{Out}_1, \text{Out}_2, \dots \} \\ \text{b. H-eval}(\text{Out}_i, 1 \leq i \leq \infty) & \rightarrow \text{Out}_{\text{real}} \end{array}$$

The grammar must define a pairing of underlying and surface forms, ( $\text{input}_i, \text{output}_j$ ). Each input is associated with a candidate set of possible analyses by the function Gen (short for 'generator'), a fixed part of Universal Grammar. In the rich representational system employed below, an output form retains its input as a subrepresentation, so that departures from faithfulness may be detected

---

<sup>2</sup> This kind of reasoning is familiar at the level of grammar selection in the form of the Evaluation Metric (Chomsky 1951, 1965). On this view, the resources of UG define many grammars that generate the same language; the members of that set are evaluated, and the optimal grammar is the real one.

by scrutiny of output forms alone. A ‘candidate’ is an input-output pair, here formally encoded in what is called ‘Out<sub>i</sub>’ in (2). Gen contains information about the representational primitives and their universally irrevocable relations: for example, that the node  $\sigma$  may dominate a node *Onset* or a node  $\mu$  (implementing some theory of syllable structure), but never *vice versa*. Gen will also determine such matters as whether every segment must be syllabified – we assume not, below, following McCarthy 1979 *et seq.* – and whether every node of syllable structure must dominate segmental material – again, we will assume not, following Itô 1986, 1989. The function H-eval determines the relative Harmony of the candidates, imposing an order on the entire set. An optimal output is at the top of the harmonic order on the candidate set; by definition, it best satisfies the constraint system. Though Gen has a role to play, the burden of explanation falls principally on the function H-eval, a construction built from well-formedness constraints, and the account of interlinguistic differences is entirely tied to the different ways the constraint-system H-eval can be put together, given UG.

H-eval must be constructible in a general way if the theory is to be worth pursuing. There are really two notions of generality involved here: general with respect to UG, and therefore cross-linguistically; and general with respect to the language at hand, and therefore across constructions, categories, descriptive generalizations, etc. These are logically independent, and success along either dimension of generality would count as an argument in favor of the optimality approach. But the strongest argument, the one that is most consonant with the work in the area, and the one that will be pursued here, broaches the distinction, seeking a formulation of H-eval that is built from maximally universal constraints which apply with maximal breadth over an entire language.

Optimality Theory, in common with much recent work, shifts the burden from the theory of operations (Gen) to the theory of well-formedness (H-eval). To the degree that the theory of well-formedness can be put generally, the theory will fulfill the basic goals of generative grammar. To the extent that operation-based theories cannot be so put, they must be rejected.

Among possible developments of the optimality idea, we need to distinguish some basic architectural variants. Perhaps nearest to the familiar derivational conceptions of grammar is what we might call ‘harmonic serialism’, by which Gen provides a set of candidate analyses for an input, which are harmonically evaluated; the optimal form is then fed back into Gen, which produces another set of analyses, which are then evaluated; and so on until no further improvement in representational Harmony is possible. Here Gen might mean: ‘do any *one* thing: advance all candidates which differ in one respect from the input.’ The Gen  $\rightleftharpoons$  H-eval loop would iterate until there was nothing left to be done or, better, until nothing that could be done would result in increased Harmony. A significant proposal of roughly this character is the *Theory of Constraints and Repair Strategies* of Paradis 1988ab, with a couple of caveats: the *constraints* involved are a set of parochial level-true phonotactic statements, rather than being universal and violable, as we insist; and the *repair strategies* are quite narrowly defined in terms of structural description and structural change rather than being of the ‘do-onto- $\alpha$ ’ variety. A key aspect of Paradis’s work is that it confronts the problem of well-definition of the notion ‘repair’: what to do when applying a repair strategy to satisfy one constraint results in violation of another constraint (at an intermediate level of derivation). Paradis refers to such situations as ‘constraint conflicts’ and although these are not conflicts in our sense of the term — they cannot be, since all of her constraints are surface- or level-true and therefore never disagree among themselves in the assessment of output well-formedness — her work is of unique importance in addressing and shedding light on fundamental complexities in

the idea of wellformedness-driven rule-application. The ‘persistent rule’ theory of Myers 1991 can similarly be related to the notion of Harmony-governed serialism. The program for *Harmonic Phonology* in Goldsmith 1990, 1991 is even more strongly of this character; within its lexical levels, all rules are constrained to apply harmonically. Here again, however, the rules are conceived of as being pretty much of the familiar sort, *triggered* if they increase Harmony, and Harmony itself is to be defined in specifically phonotactic terms. A subtheory which is very much in the mold of harmonic serialism, using a general procedure to produce candidates, is the ‘Move-x’ theory of rhythmic adjustment (Prince 1983, Hayes 1991/1995).<sup>3</sup>

A contrasting view would hold that the Input → Output map has no internal structure: all possible variants are produced by Gen in one step and evaluated in parallel. In the course of this work, we will see instances of both kinds of analysis, though we will focus predominantly on developing the parallel idea, finding strong support for it, as do McCarthy & Prince 1993. Definitive adjudication between parallel and serial conceptions, not to mention hybrids of various kinds, is a challenge of considerable subtlety, as indeed the debate over the necessity of serial Move- $\alpha$  illustrates plentifully (e.g. Aoun 1986, Browning 1991, Chomsky 1981), and the matter can be sensibly addressed only after much well-founded analytical work and theoretical exploration.

Optimality Theory abandons two key presuppositions of earlier work. First, that it is possible for a grammar to narrowly and parochially specify the Structural Description and Structural Change of rules. In place of this is Gen, which generates for any given input a large space of candidate analyses by freely exercising the basic structural resources of the representational theory. The idea is that the desired output lies somewhere in this space, and the constraint system of the grammar is strong enough to single it out. Second, Optimality Theory abandons the widely held view that constraints are language-particular statements of phonotactic truth. In its place is the assertion that constraints are essentially universal and of very general formulation, with great potential for disagreement over the well-formedness of analyses; an individual grammar consists of a ranking of these constraints, which resolves any conflict in favor of the higher-ranked constraint. The constraints provided by Universal Grammar are simple and general; interlinguistic differences arise from the permutations of constraint-ranking; typology is the study of the range of systems that re-ranking permits. Because they are ranked, constraints are regularly violated in the grammatical forms of a language. Violability has significant consequences not only for the mechanics of description, but also for the process of theory construction: a new class of predicates becomes usable in the formal theory, with a concomitant shift in what we can think the actual generalizations are. We cannot expect the world to stay the same when we change our way of describing it.

---

<sup>3</sup> An interesting variant is what we might call ‘anharmonic serialism’, in which Gen produces the candidate set by a nondeterministic sequence of constrained procedures (‘do one thing; do another one’) which are themselves not subject to harmonic evaluation. The candidate set is derived by running through every possible sequence of such actions; harmonic evaluation looks at this candidate set. To a large extent, classical Move- $\alpha$  theories (Chomsky, 1981) work like this.

### 1.3 Overall Structure of the Argument

This work falls into three parts. Part I develops the basic groundwork, theoretical and empirical, and illustrates the characteristic kinds of analytical results that can be gotten from the theory. Part II propounds a theory of universal syllable typology at two levels of idealization, drawing on and then advancing beyond various constraints introduced in Part I. The syllable structure typology provides the basis for a full-scale analysis of the rich system of prosodically-conditioned alternations in the Lardil nominal paradigm. Part III begins with an investigation of the way that inventories are delimited both in UG and in particular grammars. A variety of issues are then explored which have to do with the conceptual structure of the theory and with its relation to other work along the same general lines. We conclude with an Appendix containing proofs of some theorems stated in the text proper and other material of interest.

The argument ranges over a variety of issues, problems, generalizations, and theoretical constructions. Some are treated rapidly, with the aim of extracting a general point, others are pursued in detail; sometimes the treatment is informal, at other times it is necessary to formalize carefully so that nonobvious results can be established by explicit proof. We have tried to segregate and modularize as much as possible, but the reader should feel free on first reading to tunnel through bits that do not appeal: the formalist can surely find another formal patch up ahead, the connoisseur of generalizations another generalization. We have tried to sign-post the way in the text.

If the reader's interest is piqued by the present contents, the following works, which make use of Optimality Theory in various ways, may be of interest:

- Archangeli, Diana and Douglas Pulleyblank. 1992. *Grounded phonology*. Ms. University of Arizona and University of British Columbia.
- Black, H. Andrew. 1993. Constraint-ranked derivation: truncation and stem binarity in Southeastern Tepehuan. Ms. UC Santa Cruz.
- Churchyard, Henry. 1991. Biblical Hebrew prosodic structure as the result of preference-ranked constraints.
- Goodman, Beverley. 1993. The integration of hierarchical features into a phonological system. Doctoral dissertation, Cornell University.
- Hung, Henrietta. 1992. Relativized suffixation in Choctaw: a constraint-based analysis of the verb grade system. Ms. Brandeis University.
- Hung, Henrietta. in preparation. *The rhythmic and prosodic organization of edge constituents*. Doctoral dissertation. Brandeis University.
- Itô, Junko, Yoshihisa Kitagawa, and R. Armin Mester. 1992. prosodic type preservation in Japanese: evidence from *zuijya-go*. SRC-92-05. Syntax Research Center. UC Santa Cruz.
- Itô, Junko and R. Armin Mester. 1992. Weak layering and word binarity. Ms. University of California. Santa Cruz.
- Itô, Junko and R. Armin Mester. to appear. Licensed segments and safe paths. In *Constraints, violations, and repairs in phonology. Special issue of the Canadian Journal of Linguistics*.
- Kirchner, Robert. 1992a. *Harmonic Phonology within One Language: An Analysis of Yidin'*. MA thesis. University of Maryland. College Park.
- Kirchner, Robert. 1992b. Yidin' prosody in Harmony Theoretic Phonology. Ms. UCLA.

- Legendre, Géraldine, William Raymond, and Paul Smolensky. Analytic typology of case marking and grammatical voice.
- McCarthy, John. to appear. A case of surface constraint violation. *Canadian Journal of Linguistics*. special issued edited by Carole Paradis, Darlene LaCharité, and Emmanuel Nikiema.
- McCarthy, John and Alan Prince. 1993. *Prosodic Morphology I: constraint interaction and satisfaction*.
- Mester, R. Armin. to appear. The quantitative trochee in Latin. *Natural Language & Linguistic Theory*.
- Prince, Alan. 1990/92. Quantitative consequences of rhythmic organization.
- Rosenthal, Sam. in preparation. *The phonology of vowels and glides*. Doctoral dissertation. University of Massachusetts, Amherst.
- Samek-Lodovici, Vieri. 1992. Universal constraints and morphological gemination: a crosslinguistic study. Ms. Brandeis University.
- Samek-Lodovici, Vieri. 1993. A unified analysis of crosslinguistic morphological gemination. In *Proceedings of CONSOL-1*.
- Selkirk, Elisabeth. 1993. The prosodic structure of functional elements: affixes, clitics, and words. handout of talk presented at Signal to Syntax Conference. Brown University.
- Sherer, Tim. in preparation. *Prosodic Phonotactics*. Doctoral dissertation. University of Massachusetts. Amherst.
- Yip, Moira. 1992. Cantonese loan word phonology and Optimality Theory. To appear in *Journal of East Asian Linguistics*.
- Yip, Moira. 1993. Phonological constraints, optimality, and phonetic realization in Cantonese. UPenn Colloquium.
- Zec, Draga. in preparation. Coda constraints and conditions on syllable weight. Cornell University.
- Zoll, Cheryl. 1992. When syllables collide: a theory of alternating quantity. Ms. Brandeis University.
- Zoll, Cheryl. 1993. Ghost consonants and optimality. WCCFL, Santa Cruz.

# PART I

## Optimality and Constraint Interaction

## Overview of Part I. §§2-5.

Our first goal will be to establish that the notion of optimality is, as claimed, indispensable to grammar. In §2 we will argue this point from the results of Dell and Elmedlaoui in their landmark study of syllabification in Imdlawn Tashlhiyt Berber. In the course of this argument, we will introduce the notion of constraint domination and the fundamental mechanism for computing optimality with respect to a set of constraints that have been prioritized with this notion. We then move on in §3 and §4 to analyze fundamental recurrent patterns of grammatical generalization, showing that constraint domination explicates and improves on the notions of *triggering* and *blocking* that figure prominently in current linguistic discussion. We examine a number of phenomena central to prosodic theory, including (aspects of) the relation between foot structure and syllable structure; the interactions of prominence, minimality, and extrametricality; and the relation between syllable structure and the prosodic-morphological processes of edge-oriented infixation, arguing that proper understanding of constraint domination sheds new light on these phenomena. The formal theory of extrametricality is dissolved into interaction effects between independently-required constraints. We conclude §4 with an analysis of prosodic structure in Latin which brings together the various empirical and theoretical themes pursued in the discussion. Part I draws to a close with a formal characterization of the notion ‘evaluation with respect to a constraint hierarchy’ and study of some properties of constraint ranking.



## 2. Optimality in Grammar: Core Syllabification in Imdlawn Tashlhiyt Berber

Here we argue that certain grammatical processes can only be properly understood as selecting the *optimal output* from among a set of possibilities, where the notion *optimal* is defined in terms of the constraints bearing on the grammatical domain at issue.

### 2.1 The Heart of Dell & Elmedlaoui

The Imdlawn Tashlhiyt dialect of Berber (ITB) has been the object of a series of remarkable studies by François Dell and Mohamed Elmedlaoui (Dell & Elmedlaoui 1985, 1988, 1989). Perhaps their most surprising empirical finding is that in this language any segment — consonant or vowel, obstruent or sonorant — can form the nucleus of syllable. One regularly encounters syllables of the shape *tK*, *rB*, *xZ*, *wL*, for example. (Capitalization represents nucleus-hood of consonants.) The following table provides illustrative examples, with periods used to mark syllable edges:<sup>4</sup>

Nucleus Type	Example	Morphology	Reference
voiceless stop	.ra.t <b>K</b> .ti.	ra-t-kti	1985: 113
voiced stop	.b <b>D</b> .dL. .ma.ra.t <b>G</b> t.	bddl ma=ra-t-g-t	1988: 1 1985: 113
voiceless fricative	.t <b>F</b> .t <b>K</b> t. .t <b>X</b> .z <b>N</b> t.	t-ftk-t t-xzn-t	1985: 113 1985: 106
voiced fricative	.t <b>xZ</b> .nakk <sup>w</sup> .	t-xzn#nakk <sup>w</sup>	1985: 113
nasal	.tz <b>M</b> t. .t <b>M</b> .z <b>h</b> .	t-zmt t-mz <b>h</b>	1985: 112 1985: 112
liquid	.t <b>R</b> .g <b>L</b> t.	t-rgl-t	1985: 106
high vowel	.i <b>l</b> .d <b>i</b> . .ra.t.l <b>u</b> l.t.	i-ldi ra-t-lul-t	1985: 106 1985: 108
low vowel	.t <b>R</b> .b <b>a</b> .	t-rba	1985: 106

Dell and Elmedlaoui marshal a compelling range of evidence in support of the claimed patterns of syllabification. In addition to native speaker intuition, they adduce effects from segmental phonology (emphasis spread), intonation, versification practice, and prosodic morphology, all of which agree in respecting their syllabic analysis.

<sup>4</sup> Glosses are *ratkti* ‘she will remember’; *bddl* ‘exchange!’; *maratgt* ‘what will happen to you?’; *tftkt* ‘you (2psg) suffered (pf.) a strain’; *txznt* ‘you stored’; *txznakk<sup>w</sup>* ‘she even stockpiled’; *tzmt* ‘it(f.) is stifling’; *tmz**h*** ‘she jested’; *trgl* ‘you locked’; *ildi* ‘he pulled’; *ratlult* ‘you will be born’; *trba* ‘she carried-on-her-back’.

The domain of syllabification is the phonological phrase. All syllables must have onsets except when they occur in absolute phrase-initial position. There, syllables may begin with vowels, either with or without glottal striction (Dell & Elmedlaoui 1985: 127 fn. 20), evidently a matter of phonetic implementation. Since any segment at all can form the nucleus of a syllable, there is massive potential ambiguity in syllabification, and even when the onset requirement is satisfied, a number of distinct syllabifications will often be potentially available. But the actual syllabification of any given string is almost always unique. Dell & Elmedlaoui discovered that assignment of nuclear status is determined by the relative sonority of the elements in the string. Thus we find the following typical contrasts:

### (3) Sonority Effects on Nuclear Status

- (a)  $tZMt$  —  $*tZmt$  ‘*m* beats *z* as a nucleus’  
 (b)  $rat.lult$  —  $*ra.tL.wLt$ . ‘*u* beats *l* as a nucleus’

*Orthography*: we write *u* for the nuclear version, *w* for the marginal version of the high back vocoid, and similarly for *i* and *y*: as with every other margin/nucleus pair, we assume featural identity.

All the structures in (3), including the ill-formed ones, are locally well-formed, composed of licit substructures. In particular, there is nothing wrong with syllables  $tZ$ ,  $tL$ , or  $wLt$  nor with word-final sequences  $mt$  — but the more sonorous nucleus is chosen in each case. By examining the full range of such contrasts, Dell and Elmedlaoui establish the relevance of the following familiar kind of 8-point hierarchy:

### (4) Sonority Scale

|Low V|>|High V|>|Liquid|>|Nasal|>|Voiced Fric.|>|Voiceless Fric.|>|Voiced Stop|>|Voiceless Stop|

We write  $|\alpha|$  for the sonority or intrinsic prominence of  $\alpha$ .

With the sonority scale in hand, Dell and Elmedlaoui then propose an iterative syllable-construction procedure that is designed to select the correct nuclei. Their algorithm can be stated in the following way, modified slightly from Dell & Elmedlaoui 1985: 111(15):

### (5) Dell–Elmedlaoui Algorithm for Core Syllabification (DEA)

Build a core syllable (“CV”) over each substring of the form XY, where

X is any segment (except [a]), and

Y is a matrix of features describing a step of the sonority scale.

Start Y at the top of the sonority scale and replace it successively with the matrix of features appropriate to the next lower step of the scale.

(Iterate from Left to Right for each fixing of the nuclear variable Y.)

Like all such procedures, the DEA is subject to the Free Element Condition (FEC: Prince 1986), which holds that rules establishing a level of prosodic structure apply only to elements that are not already supplied with the relevant structure. By the FEC, the positions analyzed by the terms X,Y must be free of syllabic affiliation. Effectively, this means that any element seized as an onset is no

longer eligible to be a nucleus, and that a segment recruited to nucleate a syllable is not then available to serve as an onset.

There are other syllabification phenomena in ITB that require additional rules beyond the DEA; we will abstract away from these and focus on the sense of DEA itself.<sup>5</sup> We will also put aside some wrinkles in the DEA which are related to parenthesized expressions in (5) — the lack of a glide counterpart for /a/, the phrase-initial loosening of the onset requirement, and the claimed left-to-rightness of the procedure.<sup>6</sup>

The DEA is a rule, or rather a schema for rules, of exactly the classical type  $A \rightarrow B / C \text{---} D$ . Each rule generated by the schema has a Structural Description specified in featural terms and a Structural Change ('construct a core syllable'). To see how it works, consider the following derivations:

---

<sup>5</sup> Not the least of these is that syllables can have codas; the DEA serves essentially to locate syllable nuclei, which requires that onsets be taken into consideration. But it is not difficult to imagine plausible extensions which lead to adjunction of codas. More subtle, perhaps, are these phenomena:

- a. obstruents are always nonsyllabic in the envs. #— and —#.
- b. sonorant C's are optionally nonsyllabic —# under certain conditions.
- c. the 1st element of a tautomorphic geminate is never an onset.

In addition, the DEA does completely resolve sequences /~aa~/, which according to other sources, surface as ~aya~ (Guerssel 1985). The appropriate approach to epenthetic structure within OT involves the constraint FILL, which makes its appearance below in §3.1 and receives full discussion in §6.

<sup>6</sup> We deal with the fact that [a] cannot occupy syllable margins in §8.1.1. The commonly encountered relaxation of the onset requirement in initial position is resolved in McCarthy & Prince 1993 in terms of constraint interaction, preserving the generality of ONS. Dell & Elmedlaoui are themselves somewhat ambivalent about the need for directionality (Dell & Elmedlaoui 1985: 108); they suggest that "the requirement [of directionality] is not concerned with left to right ordering *per se*, but rather with favoring application of [the DEA] that maximize the sonority differences between [onset and nucleus]" (Dell & Elmedlaoui 1985:127, fn. 22). In addition, they note that directionality falsely predicts \*.*i.tBd.rin*. from /i=t-!bdri-n/ 'for the cockroaches', whereas the only licit syllabification is .*it.bD.rin*. The reason for this syllabification is not understood. A directionless theory leaves such cases open for further principles to decide.

## (6) DEA in Action

Steps of the DEA	/ratlult/ ‘you will be born’
<i>Seek</i> [X][+low, -cns] & <i>Build</i>	( <b>ra</b> )tlult
<i>Seek</i> [X][-low, -cns] & <i>Build</i>	(ra)t( <b>lu</b> )lt
<i>Seek</i> [X][+cns,+son,-nas]	-blocked by FEC-
<i>Seek</i> [X][+cns,+son,+nas]	—
<i>Seek</i> [X][-son,+cnt,+voi]	—
<i>Seek</i> [X][-son,+cnt,-voi]	—
<i>Seek</i> [X][-son,-cnt,+voi]	—
<i>Seek</i> [X][-son,-cnt,-voi] & <i>Build</i>	(ra)t(lu)( <b>IT</b> ) <sup>7</sup>

## (7) DEA in Action

Steps of the DEA	/txznt/ ‘you sg.stored’
<i>Seek</i> [X][+low,-cns]	—
<i>Seek</i> [X][-low,-cns]	—
<i>Seek</i> [X][+cns,+son,-nas]	—
<i>Seek</i> [X][+cns,+son,+nas] & <i>Build</i>	tx( <b>zN</b> )t
<i>Seek</i> [X][-son,+cnt,+voi]	—
<i>Seek</i> [X][-son,+cnt,-voi] & <i>Build</i>	( <b>tX</b> )(zN)t
<i>Seek</i> [X][-son,-cnt,+voi]	—
<i>Seek</i> [X][-son,-cnt,-voi]	—

---

<sup>7</sup> We show the form predicted by the DEA. The form is actually pronounced **rat.lu**lt. because obstruents cannot be nuclear next to phrase boundaries, as mentioned in fn. 5.

## (8) DEA in action

Steps of the DEA	/txznas/ ‘she stored for him’
<i>Seek</i> [X][+low, -cns] & <i>Build</i>	txz( <b>na</b> )s
<i>Seek</i> [X][-low, -cns]	—
<i>Seek</i> [X][+cns, +son, -nas]	—
<i>Seek</i> [X][+cns, +son, +nas]	-blocked by FEC-
<i>Seek</i> [X][-son, +cnt, +voi] & <i>Build</i>	t( <b>xZ</b> )(na)s
<i>Seek</i> [X][-son, +cnt, -voi]	-blocked by FEC-
<i>Seek</i> [X][-son, -cnt, +voi]	—
<i>Seek</i> [X][-son, -cnt, -voi]	-blocked by FEC-

The DEA provides an elegant and straightforward account of the selection of syllable nuclei in the language. But it suffers from the formal arbitrariness characteristic of re-writing rules when they are put to the task of dealing locally with problems that fall under general principles, particularly principles of output shape. (By ‘formal arbitrariness’, we mean that a formal system rich enough to allow expression of the desired rule will also allow expression of many undesired variations of the rule, so that the rule itself appears to be an arbitrary random choice among the universe of possibilities.) The key to the success of the DEA is the way that the variable Y scans the input, starting at the top of the sonority scale and descending it step by step as the iterative process unfolds. We must ask, why start at the top? why *descend* the scale? why not use it in some more elaborate or context-dependent fashion? why apply the scale to the nucleus rather than the onset? <sup>8</sup>

The answers are to be found in the theory of syllable structure markedness, which is part of Universal Grammar. The more sonorous a segment is, the more satisfactory it is as a nucleus. Conversely, a nucleus is more satisfactory to the degree that it contains a more sonorous segment. It is clear that the DEA is designed to produce syllables with optimal nuclei; to ensure that the syllables it forms are the most *harmonic* that are available, to use the term introduced in §1. Dell and Elmedlaoui clearly understand the role of sonority in choosing between competing analyses of a given input string; they write:

When a string ...PQ... could conceivably be syllabified as ...Pq... or as ...pQ... (i.e. when either syllabification would involve only syllable types which, when taken individually, are possible in ITB), the only syllabification allowed by ITB is the one that takes as a syllabic peak the more sonorous of the two segments.

— Dell & Elmedlaoui 1985:109

---

<sup>8</sup> These are exactly the sort of questions that were fruitfully asked, for example, of the classic TG rule of Passive that moved subject and object, inserted auxiliaries, and formed a PP: why does the post-verbal NP move *up* not *down*? why does the subject NP move at all? why is by+NP a PP located in a PP position? and so on.

But if phonology is couched in re-writing rules, this insight cannot be cashed in as part of the function that assigns structural analyses. It remains formally inert. Dell and Elmedlaoui refer to it as an “empirical observation,” emphasizing its extra-grammatical status.

The DEA itself makes no contact with any principles of well-formedness; it merely scans the input for certain specific configurations, and acts when it finds them. That it descends the sonority scale, for example, can have no formal explanation. But the insight behind the DEA can be made active if we re-conceive the process of syllabification as one of choosing the optimal output from among the possible analyses rather than algorithmic structure-building. Let us first suppose, with Dell and Elmedlaoui, that the process of syllabification is serial, affecting one syllable at a time (thus, that it operates like Move- $\alpha$  or more exactly, Move- $x$  of grid theory). At each stage of the process, let all possible single syllabic augmentations of the input be presented for evaluation. This set of candidates is evaluated by principles of syllable well-formedness and the most harmonic structure in the set is selected as the output. We can state the process informally as follows:

**(9) Serial Harmonic Syllabification (informal).**

Form the optimal syllable in the domain.

Iterate until nothing more can be done.

This approach depends directly on the principles of well-formedness which define the notion ‘optimal’. No instructions are issued to the construction process to contemplate only one featurally-specified niche of the sonority scale. Indeed, the Harmonic syllabification algorithm has no access to any information at all about absolute sonority level or the specific featural composition of vowels, which are essential to the DEA; it needs to know whether segment  $\alpha$  is *more* sonorous than segment  $\beta$ , not what their sonorities or features actually are. All possibilities are entertained simultaneously and the choice among them is made on grounds of general principle. That you start at the top of the scale, that you descend the scale rather than ascending it or touring it in some more interesting fashion, all this follows from the universal principles that define the relative Harmony of nucleus-segment pairings. The formal arbitrariness of the DEA syllable-constructing procedure disappears because the procedure itself (‘make a syllable’) has been stripped of intricacies.<sup>9</sup>

This is an instance of Harmony-increasing processing (Smolensky 1983, 1986; Goldsmith 1991, 1993). The general rubric is this:

**(10) Harmonic Processing**

Go to the most harmonic available state.

We speak not of ‘relative well-formedness’ but rather of *relative Harmony*: Harmony is a well-formedness scale along which a maximal-Harmony structure is well-formed and all other structures are ill-formed.

---

<sup>9</sup> Further development of this idea could eliminate complications at the level of the general theory; in particular, the appearance of obeying the Free Element Condition during serial building of structure could be seen to follow from the fact that disobeying it inevitably decrements the Harmony of the representation.

We conclude that the Dell-Elmedlaoui results establish clearly that harmonic processing is a grammatical mechanism, and that optimality-based analysis gives results in complex cases. Let us now establish a formal platform that can support this finding.

## 2.2 Optimality Theory

What, then, is the *optimal* syllable that Harmonic Syllabification seeks? In the core process that we are focusing on, two constraints are at play, one ensuring onsets, the other evaluating nuclei. The onset constraint can be stated like this (Itô 1986, 1989):

(11) **The Onset Constraint (ONS)**. Syllables must have onsets (except phrase initially).

As promised, we are not going to explicate the parenthesized caveat, which is not really part of the basic constraint (see McCarthy & Prince 1993: §4). The nuclear constraint looks like this:<sup>10</sup>

(12) **The Nuclear Harmony Constraint (HNUC)**. A higher sonority nucleus is more harmonic than one of lower sonority.

*I.e.* If  $|x| > |y|$  then  $\text{Nuc}/x > \text{Nuc}/y$ .

The formalizing restatement appended to the constraint uses some notation that will prove useful.

For ‘x is more harmonic than y’ we write  $x > y$ .

For ‘the intrinsic prominence of x’ we write  $|x|$ .

‘A/x’ means ‘x belongs to category A, x is the constituent-structure child of A’

The two kinds of order  $>$  and  $>$  are distinguished notationally to emphasize their conceptual distinctness. Segments of high sonority are not more harmonic than those of lower sonority. It is only when segments are contemplated in a structural context that the issue of well-formedness arises. It is necessary to specify not only the relevant constraints, but also the set of candidates to be evaluated. To do this we need to spell out the function Gen that admits to candidacy a specific range of structurings or parses of the input. In the case at hand, we want something roughly like this:

(13) **Gen** ( $input_i$ ): the set of (partial) syllabifications of  $input_i$ , which differ from  $input_i$  in no more than one syllabic adjunction.

For any form  $input_i$  to undergo Serial Harmonic Syllabification, the candidate set  $\text{Gen}(input_i)$  must be evaluated with respect to the constraints ONS and HNUC. There would be little to say if evaluation were simply a matter of choosing the candidate that satisfies both constraints. Crucially,

---

<sup>10</sup> It is also possible to conceive of the operative constraint in a kind of ‘contrapositive’ manner. Because all underlying segments of ITB are parsed, a segment is a nucleus iff it is not a member of the syllable margin. Consequently, negative constraints identifying the badness of syllable margins can have the same effect as positive constraints identifying the goodness of nuclei. We investigate this approach below in §8.1.1, §8.3.3, §8.4.2.

and typically, this straightforward approach cannot work. Conflict between the constraints ONS and HNUC is unavoidable; there are candidate sets in which no candidate satisfies both constraints.

Consider, for example, the syllabification of the form /h<sup>h</sup>aul-t<sup>n</sup>/ ‘make them (m.) plentiful’ (Dell & Elmedlaoui 1985:110). Both ONS and HNUC agree that the core syllable *h<sup>h</sup>a* should be formed: it has an onset as well as the best possible nucleus. Similarly, we must have a final syllable *t<sup>n</sup>*. But what of the rest of the string? We have two choices for the sequence /ul/: a superior nucleus lacking an onset, as in *ul*; or an onsetless syllable with an inferior nucleus, as in *wL*. This situation can be perspicuously displayed in tabular form:<sup>11</sup>

#### (14) Constraint Inconsistency

Candidates	ONS	HNUC
/h <sup>h</sup> aul-t <sup>n</sup> /		
~.wL.~		l
~.ul.~	*	u

The cells contain information about how each candidate fares on the relevant constraint. A blank cell indicates that the constraint is satisfied; a star indicates violation. (In the case of a scalar constraint like HNUC we mention the contents of the evaluated element.) The first form succeeds on ONS, while the second form violates the constraint. The relative performance is exactly the opposite on HNUC: because  $|u| > |l|$ , the second, onsetless form has the better nucleus. The actual output is, of course, *.h<sup>h</sup>a.wL.t<sup>n</sup>*. The onset requirement, in short, takes priority.

Such conflict is ubiquitous, and to deal with it, we propose that a relation of *domination*, or priority-ranking, can be specified to hold between constraints. When we say that one constraint *dominates* another, we mean that when they disagree on the relative status of a pair of candidates, the dominating constraint makes the decision. If the dominating constraint does not decide between the candidates — as when both satisfy or both violate the constraint equally — then the comparison is passed to the subordinate constraint. (In the case of a more extensive hierarchy, the same method of evaluation can be applied repeatedly.)

In the case at hand, it is clear that ONS must dominate HNUC. The top priority is to provide syllables with onsets; the relative Harmony of nuclei is a subordinate concern whose force is felt only when the ONS issue is out of the way. We will write this relation as  $ONS \gg HNUC$ . Given such a hierarchy, an optimality calculation can be usefully presented in an augmented version of display (14) that we will call a *constraint tableau*:

<sup>11</sup> Properly speaking, if we limit our attention to the core syllable stage of the procedure, we should be comparing core *.u* with core *.wL*. But the comparison remains valid even after coda consonants are adjoined and we wish to emphasize that the two cited analyses of /h<sup>h</sup>aul-t<sup>n</sup>/ differ only in treatment of the sequence /ul/.



(15) **Constraint Tableau** for partial comparison of candidates from /haultn/

Candidates	ONS	HNUC
☞ ~.wL.~		l
~.ul.~	* !	u

Constraints are arrayed across the top of the tableau in domination order. As above, constraint violations are recorded with the *mark* \*, and blankness indicates total success on the constraint. These are the theoretically important conventions; in addition, there is some clarificatory typography. The symbol ☞ draws the eye to the optimal candidate; the ! marks the *crucial* failure for each suboptimal candidate, the exact point where it loses out to other candidates. Cells that do not participate in the decision are shaded. In the case at hand, the contest is decided by the dominant constraint ONS; HNUC plays no role in the comparison of .wL. and .ul. HNUC is literally irrelevant to this particular evaluation, as a consequence of its dominated position — and to emphasize this, we shade its cells. Of course, HNUC is not irrelevant to the analysis of *every* input; but a precondition for relevance is that there be a set of candidates that tie on ONS, all passing it or all failing it to the same extent.


If we were to reverse the domination ranking of the two constraints, the predicted outcome would be changed: now .ul. would be superior to .wL. by virtue of its relative success on HNUC, and the ONS criterion would be submerged. Because of this, the ranking ONS >> HNUC is *crucial*; it must obtain in the grammar of Berber if the actual language is to be generated.

The notion of domination shows up from time to time in one form or another in the literature, sometimes informally, sometimes as a clause clarifying how a set of constraints is to be interpreted. For example, Dell and Elmedlaoui write, “The prohibition of hiatus...*overrides*” the nuclear sonority comparison (Dell & Elmedlaoui 1985: 109, emphasis added). For them, this is an extra-grammatical observation, with the real work done by the Structural Descriptions provided by the DEA and the ordering of application of the subrules. Obviously, though, the insight is clearly present. Our claim is that the notion of domination, or ‘over-riding’, is the truly fundamental one. What deserves extra-grammatical status is the machinery for constructing elaborately specific Structural Descriptions and modes of rule application.

To see how Serial Harmonic Syllabification (9) proceeds, let us examine the first stage of syllabifying the input /txznt/ ‘you sg. stored, pf.’. It is evident that the first syllable constructed must be .zN. — it has an onset, and has the highest sonority nucleus available, so no competing candidate can surpass or even equal it. A more discursive examination of possibilities might be valuable; the larger-scale comparisons are laid out in the constraint tableau below.

Here are (some of the) leading candidates in the first round of the process:


(16) **Constraint Tableau for Serial Syllabification** of /txznt/ (partial, first step)

Candidates	ONS	HNUC	Comments
 tx(zN)t		n	optimal: onsetted, best available nucleus
txz(N)t	* !	n	no onset, HNUC irrelevant
t(xZ)nt		z !	z  <  n
(tX)znt		x !	x  <  n
txz(nT)		t !	t  <  n

Syllabic parsing is conceived here as a step-by-step serial process, just as in the DEA. A candidate set is generated, each produced by a single licit change from the input; the relative status of the candidates is evaluated, yielding an optimal candidate (the output of the first step); and that output will then be subject to a variety of further single changes, generating a new candidate set to be evaluated; and so on, until there are no bettering changes to be made: the final output has then been determined.

This step-by-step Harmony evaluation is not intrinsic to the method of evaluation, though, and, in the more general context, when we discard the restricted definition of Gen in (13), it proves necessary to extend the procedure so that it is capable of evaluating entire parsed strings, and not just single (new) units of analysis. To do this, we apply the same sort of reasoning used to define domination, but *within* the constraint categories. To proceed by example, consider the analysis of /txznt/ taking for candidates all syllabified strings. We present a sampling of the candidate space.

(17) **Parallel Analysis of Complete Syllabification** of /txznt/

Candidates	ONS	HNUC	Comments
 .tX.zNt.		n x	optimal
.Tx.zNt.		n t !	n  =  n ,  t  <  x
.tXz.nT.		x ! t	x  <  n , t irrelevant
.txZ.Nt.	* !	n z	HNUC irrelevant
.T.X.Z.N.T.	* ! ***	n z x t t	HNUC irrelevant

In evaluating the candidates we have kept to the specific assumptions mentioned above: the onset requirement is suspended phrase-initially, and the nonnuclear status of peripheral obstruents is, as in the DEA itself, put aside (see fn. 5).

In this tableau, all the relevant information for harmonic evaluation of the parse of the whole string is present. We start by examining the first column, corresponding to the dominant constraint ONS. Only the candidates which fare best on this constraint survive for further consideration. The first three candidates all have syllables with onsets; the last two do not (to varying degrees). Lack of onset in even a single non-initial syllable is immediately fatal, because of the competing candidates which satisfy ONS.

The remaining three parses are not distinguished by ONS, and so HNUC, the next constraint down the hierarchy, becomes relevant. These three parses are compared by HNUC as follows. The most sonorous nucleus of each parse is examined: these are the most harmonic nuclei according to HNUC. For each of the first two candidates the most sonorous nucleus is *n*. For the last candidate, the most sonorous nucleus is *x*, and it drops out of the competition since *n* is more sonorous than *x*. We are left with the first two candidates, so far tied on all comparisons. The HNUC evaluation continues now to the next-most-harmonic nuclei, where the competition is finally settled in favor of the first candidate .tX.zNt.

What we have done, in essence, is to replace the iterative procedure (act/evaluate, act/evaluate,...) with a recursive scheme: collect the results of all possible actions, then sort recursively. Rather than producing and pruning a candidate set at each step of sequential processing, striving to select at each step the action which will take us eventually to the correct output, the whole set of possible parses is defined and harmonically evaluated. The correct output is the candidate whose complete structure best satisfies the constraint hierarchy. And ‘best satisfies’ can be recursively defined by descending the hierarchy, discarding all but the best possibilities according to each constraint before moving on to consider lower-ranked constraints.

The great majority of analyses presented here will use the parallel method of evaluation. A distinctive prediction of the parallel approach is that there can be significant interactions of the top-down variety between aspects of structure that are present in the final parse. In §3, §4 and §7 we will see a number of cases where this is borne out, so that parallelism is demonstrably crucial; further evidence is presented in McCarthy & Prince 1993. ‘Harmonic serialism’ is worthy of exploration as well, and many hybrid theories can and should be imagined; but we will have little more to say about it. (But see fn. 49 below on Berber syllabification.)

The notion of parallel analysis of complete parses in the discussion of constraint tableau (17) is the crucial technical idea on which many of our arguments will rest. It is a means for determining the relative harmonies of entire candidate parses from a set of conflicting constraints. This technique has some subtleties, and is subject to a number of variant developments, so it is worth setting out with some formal precision exactly what we have in mind. A certain level of complexity arises because there are two dimensions of structure to keep track of. On the one hand, each individual constraint typically applies to several substructures in any complete parse, generating a *set* of evaluations. (ONS, for example, examines every syllable, and there are often several of them to examine.) On the other hand, every grammar has multiple constraints, generating multiple sets of evaluations. Regulating the way these two dimensions of multiplicity interact is a key theoretical commitment.

Our proposal is that evaluation proceeds by constraint. In the case of the mini-grammar of ONS and HNUC, entire syllabifications are first compared via ONS alone, which examines each syllable for an onset; should this fail to decide the matter, the entire syllabifications are compared via HNUC alone, which examines each syllable’s nucleus.

Another way to use the two constraints would be to examine each (completely parsed) candidate syllable-by-syllable, assessing each syllable on the basis of the syllabic mini-grammar. The fact that ONS dominates HNUC would then manifest itself in the Harmony assessment of each individual syllable. This is also the approach most closely tied to continuous Harmony evaluation during a step-by-step constructive derivation. Here again, we do not wish to dismiss this conception, which is surely worthy of development. Crucially, however, this is not how Harmony evaluation works in the present conception (see §5.2.3.1 for further discussion).

In order to characterize harmonic comparison of candidate parses with full generality and clarity, we need to specify two things: first, a means of comparing entire candidates on the basis of a single constraint; then, a means of combining the evaluation of these constraints. The result is a general definition of *Harmonic Ordering of Forms*; this is, in its formal essence, our theory of constraint interaction in generative grammar. It is the main topic of §5.

### 2.3 Summary of discussion to date

The core syllabification of Imdlawn Tashlhiyt Berber provides a particularly clear case where the function assigning structural analyses must be based on the optimality of the output if it is to be properly founded on principle. Once the relevant principles have been moved into grammatical theory, it becomes possible to undertake a radical simplification of the generative procedure that admits candidate syllable structures. The focus shifts away from the effort to construct an algorithm that assembles the correct structure piece-by-piece, an effort that we believe is doomed to severe explanatory shortcomings. Linguistic theory, properly conceived, simply has little to say about such constructional algorithms, which (we claim) are no more than implementations of grammatical results in a particular computational-like framework. The main explanatory burden falls the constraints themselves, and on the apparatus that governs their interactions.

The Berber situation is particularly interesting in that core syllabification simply cannot proceed without the intervention of *two* distinct constraints. As with other forms of prosodic organization, the most common picture is one in which the structure is built (more-or-less) bottom-up, step-by-single-step, with each step falling under the constraints appropriate to it. Taking this seriously in the syllable structure domain, this would mean, following Levin [Blevins] (1985) and ultimately Kahn (1976), that you first locate the nuclei — the heads of syllables; then project higher order structure that includes the onsets; then project the structure that includes postnuclear consonantism. In ITB, however, as in many other languages, the availability of nuclei depends on the choice of onsets: an early step in the derivational constructive procedure, working on a low level in the structural hierarchy, depends on later steps that deal with the higher levels. Indeed, the higher level constraint is very much the more forceful. Technical solutions to this conundrum can be found in individual cases — Dell and Elmedlaoui's being a particularly clever one; but the theme will reappear persistently in every domain of prosody, defying a uniform treatment in constructionist terms.

In the theory advocated here, where outputs are evaluated, we expect exactly this kind of interaction. The whole output is open to inspection; how we choose to inspect it, or how we are forced by UG to inspect it, is not determined by the course that would be taken in bottom-up construction. The potential force of a constraint is not indexed to the level of structure that it pertains to, and under certain circumstances (UG permitting or demanding), constraint domination will be invoked to give higher-level constraints absolute priority over those relevant to the design of lower structural levels.

### 3. Generalization-Forms in Domination Hierarchies I

#### Blocking and Triggering: Profuseness and Economy

Two patterns of constraint interaction appear repeatedly in modern grammatical work. The first can be informally characterized as *Do Something Only When Necessary*. Under this rubric, a process of wide formal generality, say  $\emptyset \rightarrow V$ , nevertheless applies only in special circumstances, namely those in which it is necessary for well-formedness. The default is *not to do*; the process is *triggered* by the constraint or constraints that it subserves. In operational phonology, a ‘repair strategy’ is exactly a process which applies only when it leads to satisfaction of otherwise-violated constraints (Singh 1987, Paradis 1988ab). Epenthesis is a typical example; it is so closely tied to syllable-structure and word-structure constraints that the process itself can be given an entirely general formulation (e.g.  $\emptyset \rightarrow V$ ), with no limiting environmental restrictions, so long as it is understood to apply *only when needed*. (This notion appears in Sommerstein 1972 and is closely related to the proposals of Kisseberth 1970ab, 1972). There is a considerable body of work on the idea, including Singh, Paradis, Prunet, Myers, McCarthy, Itô, Mester, Yip, Goldsmith, though we believe it is fair to say that no consensus has emerged on how the rule-constraint interaction is to be understood; that is, how and when the rule is to be triggered by the constraint(s). In syntax the notion *Do Something Only When Necessary* appears under the heading of ‘movement as a last resort’ or, more generally, ‘Economy of Derivation’ (Chomsky 1989).

The second common pattern can be characterized as *Do Something Except When Banned*. This is the original Rossian and Kisseberthian style of rule/constraint interaction, in which the constraint *blocks* the rule: front a wh-phrase *except when* such & such conditions prevail. Here the default is *to do*; the process is inhibited under narrowly specifiable conditions. In operational phonology, deletion rules are often stated this way: delete a certain vowel everywhere, *except when* it leads to ill-formed syllable-structure, stress-clash, or violations of the OCP (Kisseberth 1970ab, 1972; Hammond 1984; McCarthy 1986). This phenomenon we might call ‘Profuseness of Derivation’.

Both patterns emerge from strict domination, where the basic fact of life is that a higher-ranked constraint forces violations of a lower-ranked constraint. How this looks to the operational describer depends on the character of the lower-ranked constraint. The economy or triggering class (*Do Something Only When*) emerges when the lower-ranked constraint bans some structural option; when the dominating constraint is at stake, the banned option will be taken — and only then. The profuseness or blocking class (*Do Something Except When*) emerges when the lower-ranked constraint *favors* some option — perhaps by blocking its blocking by a yet-lower-ranked constraint; now the high-ranked constraint can force rejection of the otherwise-favored option.

We now turn to the analysis of a range of interactions exhibiting the descriptive characters of Economy and Profuseness. We will see that constraint-domination theory exposes the common conceptual core of a number of different-seeming phenomena, leading to a deeper understanding of some poorly-resolved issues in phonological theory and practice.

## Do Something Only When: Triggering, or The Theory of Economy

### 3.1 Epenthetic Structure

Nothing ain't worth nothing, but it's free.  
— Common Misconception

To illustrate the characteristic form of the economic interaction, let us consider a pattern of epenthesis in the phonology of (Classical) Arabic. Simplifying for purposes of exposition, let us focus entirely on C-epenthesis, which supplies a glottal stop in certain environments. From an input /al-qalam+u/ ‘the-pen (+nom.)’, we get an output *ʔalqalamu*. Arabic syllables must be of the form CV(X); epenthesis ensures that the obligatory onset is present, whatever the input configuration. However, if the rule is stated with maximal generality, context-free, as  $\emptyset \rightarrow C$ , we must ask why we don't get such forms as these, which are all in fine accord with the Arabic syllable canon:

(18) **Free Epenthesis into /al-qalamu/** (all ungrammatical)

- a. ʔalqaʔlamu
- b. ʔalqalʔamu
- c. ʔalqalamʔu
- d. ʔalqaʔlaʔmu
- e. ʔalqaʔlamuʔ
- f. ʔalqalʔamuʔ
- g. ʔalqalʔaʔmu
- h. ʔalqaʔlaʔmuʔ
- i. ʔalqalʔaʔmuʔ

Let us suppose, following Selkirk 1981, Broselow 1982, Piggott & Singh 1985, and Itô 1986, 1989 quite closely, that the site of epenthesis is an empty syllabic position that arises during the course of syllabification. The phonetic value of the epenthetic item is filled in by interpretive principles that read the output of the level of structure we are concerned with. On this view, the basic principles of syllable structure assignment already have within them all that is required to produce the correct output forms. An epenthetic structure is simply a licit syllable form that contains structure not motivated by the presence of a segment. There are no additional specialized operations of repair recapitulating aspects of parsing itself; no rule “insert Onset” or the like, since that is part of what syllabification does in the first place. The candidate set of possible analyses already contains the correct output; the constraint system will locate it. We believe this is a major insight into the way grammar works.

What is called *epenthesis*, then, is a by-product of the syllable parse, which must therefore contemplate candidates that have empty structure in them. Empty structure is avoided, obviously, as indeed is anything that would lead to structural complexity in the relationship between base and surface forms. We therefore hypothesize that there is a system of constraints that deals with sources of such complexity; call these the ‘faithfulness’ constraints. Itô’s *Prosodic Licensing*, which requires that realized segments belong to syllables, is clearly among these (Itô 1986, 1989; Goldsmith 1990; Itô & Mester 1993), as is the *Melody Integrity* of Borowsky 1986. Prosodic Licensing looks up from the segment, and more generally up from any node, to check that it has a parent, and an appropriate

one. We will call this family of constraints by the name PARSE; several members of the family will be encountered below. (Unparsed elements are denied phonetic realization, in accord with the notion of ‘Stray Erasure’ familiar from McCarthy 1979, Steriade 1982, Itô 1986, 1989.) We also need the counterpart of Prosodic Licensing that looks *down* from a given node to be sure that its immediate contents are appropriate. Let’s call this family of constraints FILL, the idea being that every node must be filled properly; the idea dates back to Emonds 1970.<sup>12</sup> In the case at hand, we are interested in syllable structure, and the constraint can be stated like this:

(19) **FILL**

Syllable positions are filled with segmental material.

Arabic unmistakably exhibits the ONS constraint, which we state as follows:

(20) **ONS**

Every syllable has an Onset.

For concreteness, let us assume that Onset is an actual node in the syllable tree; the ONS constraint looks at structure to see whether the node  $\sigma$  dominates the node Onset. This provides us with the simplest formal interpretation of FILL as banning empty nodes.<sup>13</sup>

The function Gen produces candidates that meet these requirements:

---

<sup>12</sup> PARSE and FILL are conceived here in simple constituent structure terms. This suits the straightforwardly monotonic relation we assume throughout this work between input and output, and allows clean technical development of the leading ideas. The ideas themselves, of course, admit extension to the “feature-changing” arena, which shows other kinds of breaches of faithfulness to input forms. Generally conceived, the PARSE family militates against any kind of failure of underlying material to be structurally analyzed (‘loss’) and the FILL family against any kind of structure that is not strictly based on underlying material (‘gain’).

<sup>13</sup> A more general view of the matter would see FILL as a specialized member of a broad family of constraints that ban structure altogether: \*STRUC. See Zoll (1992b) for discussion of this notion. Constraints of the \*STRUC family ensure that structure is constructed minimally: a notion useful in syntax as well as phonology, where undesirable options (move- $\alpha$ ; non-branching nonterminal nodes) typically involve extra structure. For example, the suggestion in Chomsky 1986:4 that nonbranching N\* (no specifier) is disallowed, with NP directly dominating N in such cases, would be a reflex of \*STRUC, as would Grimshaw’s proposal that extended projections are minimal in extent (Grimshaw 1993). X’-theory then reduces entirely to the principle that a syntactic category has a head of the same category,  $[X] \rightarrow \dots [X] \dots$ . 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. In the original case discussed in Chomsky 1989, we have for example a contrast between *isn’t* and *doesn’t be*; the presence of *DO* is a kind of FILL violation.

(21) **Assumed Syllable Structures**a.  $\sigma \rightarrow (\text{Ons}) \text{Nuc} (\text{Coda})$ ‘if an analysis contains a node  $\sigma$ , it must dominate Nuc and may dominate Ons and Coda.b.  $\text{Ons}, \text{Coda} \rightarrow (\text{consonant})$ 

‘if an analysis contains a node Ons or Coda, it may dominate a consonant.

c.  $\text{Nuc} \rightarrow (\text{vowel})$ 

‘if an analysis contains a node Nuc, it may dominate a vowel.

Any parse at all of the input string is admitted by Gen, so long as any structure in it accords with the prescriptions of (21).<sup>14</sup>

The constraint ONS is never violated in Arabic. FILL is violated, but only to provide onsets. This establishes the domination relation  $\text{ONS} \gg \text{FILL}$ .

The Gen function for syllable structure should admit every conceivable structure, with every conceivable array of affiliations and empty and filled nodes. In order to retain focus on the issue at hand, let’s admit every structure that allows empty onsets and empty codas, deferring till §6-8 a more ambitious investigation of syllable theory.

To establish that *ʔalqalamu* is the optimal outcome, we examine here a (sampling of) the candidate set, in which the forms are fully parsed syllabically:

(22) **Syllabification under  $\text{ONS} \gg \text{FILL}$ , fully parsed forms**

Candidates	ONS	FILL
☞ .□al.qa.la.mu.		*
.al.qa.la.mu.	*!	
.□al.qa□.la.mu.		** !
.□al.qal.□a.mu.		** !
.□al.qa□.la□.mu.		** ! *
.□al.qa□.la□.mu□.		** ! **

The symbol □ notates an empty position, a nonterminal node, daughter of  $\sigma$ . Periods mark syllable-edges.

It is evident that any form violating ONS will be excluded, since there will always be competitors that meet the constraint by virtue of empty Onset nodes. Violations of FILL are therefore compelled. What then of the competition among the FILL violators? The evaluation of candidates

<sup>14</sup> We do not urge this as the ultimate theory of syllable structure; what is important for the argument is that the distinctions that it makes will be inevitably reflected in other theories. Other assumptions, for example those in McCarthy & Prince 1986 or Itô 1986, require technical re-formulation of FILL. See Hung 1992, Kirchner 1992, Samek-Lodovici 1992, 1993, McCarthy & Prince 1993.



with respect to a constraint hierarchy is to be handled by a principle of Harmonic Ordering of Forms (HOF), mentioned in §2 and treated formally below in §5. Let us suppose that HOF works to compare two candidates by examining the set of marks that they incur. The ‘First Member’ of a collection of violation-marks for a given form is the most significant mark in its collection, where the significance of a mark is determined in the obvious way by the rank of the constraint whose violation it records. The comparison of two competitors proceeds by comparing the ‘First Member’ of each competitor’s mark collection. When there is a tie, the ‘First Members’ of each collection are thrown away, until one of the competitors has none left. By the very definition of constraint violation, we have  $\emptyset > \{*\}$ , meaning that it is better to pass a constraint than to fail it. Therefore, under this conception of evaluation, the competitor with the least number of violations is the victor in the comparison.

At this point, it might be objected that we have introduced *counting* into the reckoning of optimality, contrary to our explicit assertion above. HOF does not count, however — *not even to one*. To see this, observe that no effect can ever be arranged to occur if and only if there is exactly one violation of a certain constraint. (Why? Suppose that we altered every constraint so as to add one additional violation to every candidate it assesses. The HOF is completely unaffected, and the theory operates exactly as before, even though the candidates with exactly one violation have all changed.) HOF merely distinguishes more from less, presence from absence, the empty set from all others. In particular, HOF can never determine the absolute number of violations; that is, *count* them.<sup>15</sup> HOF deals not in quantities but in comparisons, and establishes only relative rankings, not positions on any fixed absolute scale.<sup>16</sup>

This observation leads to an important result, worthy of separate statement.

### (23) Economy Property of Optimality Theory

Banned options are available only to avoid violations of higher-ranked constraints and can only be used *minimally*.

In constraint-domination grammars, the ‘economy of derivation’ pattern emerges because violations are always minimized.

Since we have ranked ONS above FILL for one particular language, it is reasonable to ask what happens when the ranking is reversed, with FILL  $\gg$  ONS. In this situation, any violation of FILL will rule a candidate out; so the ONS constraint must simply give way when there is no consonant around to fill the Onset node. We will have e.g.  $.a. > .\square a.$ , the onsetless syllable rated as better than one that is onset-containing by virtue of empty structure. This gives us a different type of language, but still a natural language, and shows that re-ranking of constraints yields a typology of admissible systems. The set of all possible rankings generates what we can call a *factorial typology* of the

---

<sup>15</sup> There a parallel here to the analysis of Imdlawn Tashlhiyt Berber nucleus choice in the conception of §2, where evaluation never knows what features are involved, just whether a candidate nucleus is more or less harmonic than its competitors. In §§8.1.1, 8.3.3, 8.4.2 below we show how to treat the ITB system in a strictly mark-sensitive way, in full accord with the HOF as described here.

<sup>16</sup> The same is of course true of the misnamed ‘feature-counting’ evaluation metric, which was hypothesized to select the optimal *grammar* from among a candidate set of grammars.

domain to which the constraints are relevant. Below in §6 we explore a factorial typology of syllable structure, based on the interaction of FILL, PARSE, ONS, and other relevant constraints.<sup>17</sup>

### 3.2 Do Something Only When: The Failure of Bottom-up Constructionism

“... piled building supra building pon the banks for the livers by the Soangso.”  
– *Finnegans Wake*, 4.

Tongan, a language of the Pacific, shows an interaction between stress and syllable structure which raises interesting issues for prosodic theory. Poser (1985), following Churchward 1953, provides the original generative discussion. Mester (1991), improving on Poser, proposes a line of analysis that can be illuminatingly pursued within constraint domination theory.

Syllables in Tongan are always open, and all heavy syllables are stressed. Since there are no syllable-closing consonants, the weight contrast is necessarily based on the distinction between short vowels (V) and long vowels or diphthongs (VV). Main-stress falls on the penultimate mora (V) of the word. This suggests that the pattern is established by bimoraic trochees, with a right-to-left directionality.

Prosodic systems of this sort inevitably face a conflict between the demands of syllabification and the demands of foot structure. Unless other constraints impinge, any sequence CVV is optimally syllabified as .CVV., avoiding a violation of ONS. But if footing is to start immediately at the right edge of the word, forms /~CVVCV/ would have to suffer splitting of the VV sequence in two, yielding .CV.(V.CV)#. In many familiar languages, e.g. Latin (Hayes 1987), the foot-parse begins one mora over in such forms, yielding ~(.CVV).CV., where the last foot does not sit at the right edge of the word. Stress settles on the penultimate *syllable* and therefore (prominence falling on the peak thereof) on the *antepenultimate* mora. This effect has been charged to a Principle of Syllabic Integrity (e.g. Prince 1976), which holds that foot-parsing may not dissect syllables.

Tongan, however, requires strict penultimacy of main-stress, and syllabification is not allowed to stand in its way. The word /ma:ma/ ‘world’ is parsed *ma.á.ma* not \**má:ma*. Churchward

<sup>17</sup> To complete the argument about epenthesis, we must consider a larger range of candidates than we have examined so far. In particular, Gen freely admits analyses in which segments are left out of syllable structure, and these analyses must be disposed of. Such ‘underparsing’ analyses violate the constraint PARSE. As will be shown in §6, the domination relation PARSE>>FILL eliminates them. To preview, observe that underparsing can lead to satisfaction of ONS, as in output ⟨V⟩.CV. from input /VCV/. (We write ⟨X⟩ for stray X with no parent node.) With ONS out of the way, PARSE and FILL are brought into direct conflict, resolved in Arabic in favor of PARSE (nothing left out), as the following tableau illustrates:

Candidates	ONS	PARSE	FILL
☞ .□V.CV.			*
⟨V⟩.CV.		* !	

(The dotted vertical line indicates that ONS and Parse are not crucially ranked with respect to each other.) The investigation in §6 of the ONS-PARSE-FILL ranking typology resolves issues of exactly this character.

provides the useful contrast *hú:* ‘go in’ (monosyllabic) versus *hu.ú.fi* ‘open officially’ (trisyllabic), both involving the stem /hu:/. Mester observes that the *last* foot must be strictly final, with exact coincidence of foot-edge and word-edge. He proposes a rule he calls Edge-Foot which builds a foot at the right edge, disturbing syllable structure if necessary. Mester notes that although the *rule* of Edge-Foot violates Syllabic Integrity in one sense, the actual output (with resyllabification) ultimately meets the constraint, since no foot edge falls in the middle of a syllable.

This apparent conundrum vanishes when the interaction is understood in terms of constraint domination. Mester’s insight into strict edgmostness can be brought into direct confrontation with the principles of syllabification. Both syllabic and stress-pattern considerations will bear simultaneously on establishing the optimal prosodic parse. Two constraints are centrally involved: ONS, and something that does the work of Mester’s Edge-Foot. Note that along with Edge-Foot, the system requires a statement to the effect that the last foot bears main stress: Mester calls on the End-Rule (Right) to accomplish this. We suggest that the most promising line of attack on the whole problem is to generalize the notion of *End-Rule* so that it can carry the entire burden, forcing penultimate ‘breaking’ as well as final prominence. Instead of demanding that prominence fall on an element in absolute edgmost position, suppose we pursue the scalar logic made usable by the present theory and demand instead that the prominence fall on the position *nearest* to the edge, comparing across candidates. In the more familiar parsing mode of #(CVV)CV#, the foot is as near to the edge as it can be while upholding the canons of syllable form, ONS in particular. In Tongan, by contrast, the heterosyllabic parse of the vowel sequence in ~CVVCV# words successfully puts the main-stress foot as far right as can be — in absolute edge position. The actual statement of the revised End-Rule is the same as before:

(24) **EDGEMOST** position of head foot in word

The most prominent foot in the word is at the right edge.

The difference is one of interpretation. We now have the means to deal with gradient or multiple violations of the constraint, in terms of a measure of distance from the edge, say syllables. Each element that intervenes between the last foot and the edge counts as a separate violation; the economy property of the theory ensures that any such violations will be minimal in all optimal forms.

For the candidate set, let us consider only those parses that are properly bracketed. In this way Gen encodes those aspects of the Principle of Syllabic Integrity that are empirically supportable. Because syllable well-formedness gives way to the edgmostness requirement, we must have EDGEMOST >> ONS. Since EDGEMOST is the only foot constraint that dominates any syllabic criteria, it follows that VV sequences elsewhere in the word will never split, regardless of their position.

The comparison of alternatives runs like this:

(25) **Tongan Penultimacy**, from /huu+fi/ ‘open officially’

Candidates	EDGEMOST	ONS
☞ hu.(ú.fi)	∅	*
(húu).fi	σ !	

The cells in the EDGEMOST column of the tableau indicate the degree of concordance with the constraint by mentioning the string intervening between the main foot and word-edge. If the order of domination is reversed, the second candidate becomes optimal and we get the more familiar pattern of stressing.

As for the observed stressing of heavy syllables throughout the word, there are a number of descriptive options. One could assume that feet are distributed throughout the word, but that (aside from main-stress) only heavy syllable foot-heads are interpreted as especially prominent, or ‘stressed’. Or one could assume with Mester that feet are not assigned generally but only to heavy syllables. Or one could assume that it’s not footing at all, but merely the intrinsic prominence of heavy syllables that observers of the language are reporting (Prince 1983, Hayes 1991/1995). Such issues can only be decided at the level of general prosodic theory, and we leave the matter open.

It is instructive to compare the constraint-domination solution with that offered by Poser (1990) in the spirit of what we might call *Bottom-Up Constructionism*: the view that structures are literally *built by re-write rules* in a strictly bottom-up fashion, constructing layer upon layer. Poser distinguishes an initial round of syllabification, the first step of construction, from a later re-adjustment stage. At first, all VV sequences throughout the entire word are parsed as bisyllabic V.V., the idea being that moras are independent of syllables. Feet are then built on the syllables thus obtained, and main-stress is determined from the feet. When all prosodic structure has been built, there is a final round of syllabic processing (post-lexical?) in which the sequence V.V coalesces everywhere except when the second mora bears main-stress.

The analytical strategy is to build structure bottom-up and then to do such repairs as are necessary. Under this approach, the unusual word-final situation (V.V before CV#) must be extended to the entire word. This is no quirk of Poser’s particular proposal; Bottom-up Constructionism leaves little choice in the matter. If syllabification is to be done with more-or-less familiar and uniform mechanisms, then there is no way to anticipate the specific demands of main-stressing when the syllable level is being constructed. Consequently, there will be levels of derivation in which the syllabification of words is highly marked, perhaps even impossible, with heterosyllabic sequences of identical vowels. No known phenomena of Tongan, however, are sensitive to the postulated V.V syllables.

Poser’s analysis shows that Bottom-Up Constructionism is empirically inconsistent with the appealing and often-assumed idea that initial syllabification is governed by markedness considerations. By this, only the least marked options are initially available to syllabify underlying forms. Indeed, we might argue that Bottom-Up Constructionism is *antithetical* to the idea. Since the burden of establishing well-formedness falls on the later rules that function as repair strategies, it is always possible to allow great divergences from well-formedness at earlier stages. It becomes *necessary* to do so in cases like Tongan, where the generality of early rules of construction can be maintained only by letting them ignore basic markedness principles.

In the present theory, divergences from the locally unmarked state of affairs are allowed at any level of structure, but only when compelled. The Tongan ranking of EDGEMOST above syllabic Harmony promotes an alternative (V.V) that is usually suppressed, but promotes it only in those circumstances where it makes a difference. Bottom-Up Constructionism has no way to express this kind of dependency. The only relevant model of interaction is one in which two general rules happen to overlap in territory, with one re-doing part of what the other has done. Here, general coalescence corrects general syllabification. Consequently, the V.V syllabification must be portrayed as general

in Tongan, and UG must be accordingly distorted to allow it as a real option that is independent of coalescence — an intolerable conclusion.

Bottom-Up Constructionism is suited to a state of affairs in which the package of constraints relevant to syllable structure completely and universally dominates the constraint-package pertaining to foot structure: SYLL-FORM  $\gg$  FOOT-FORM. Tongan shows that this hierarchy is not, in fact, rigidly adhered to. Bimoraic trochaic footing at the end of a word necessarily entails a conflict between syllable and foot constraints in words ending Heavy-Light. Some languages resolve the conflict in favor of the syllabic constraints, using the bottom-up domination pattern; others, like Tongan, place one of the foot constraints in dominant position.<sup>18</sup>

We have classified the Tongan situation as an example of the ‘do something only when’ or triggering pattern. In the operational analysis of Mester 1991, which works bottom-up, the breaking of long penultimate vowels occurs at the level of foot-formation, “as a structure-changing imposition of the foot, violating syllable integrity” (p.2). Local resyllabification is *triggered* by the requirement of utmost finality on the main foot; breaking is therefore a last resort.<sup>19</sup> We take from Mester the idea that foot-form is what’s at stake, but we recognize no operation of breaking as a necessary part of the history of the optimal form. The broken parse is among the many alternatives that are always available to construe any string with a long vowel or VV sequence in it. (There is no ‘violation of syllable integrity’ since there is never a syllable .CVV. whose integrity could be violated.) Furthermore, we recognize no special relationship of *triggering*: there is only constraint domination. Because of the constraint ONS, which it violates, any “broken” parse is going to be suboptimal unless ONS itself is appropriately dominated. In Tongan, the dominance of EDGEMOST entails that ONS will be violated (once) in the optimal parse of certain words: those ending ~CVVCV.

Tongan provides us with our first case where it is important to have the entire structural analysis present at the time of evaluation. The imposition of prosodic requirements top-down is not an isolated peculiarity of one language, however. In prosodic systems, the main stress of a domain is often required to meet its edge-location requirements at the expense of other characteristics of the pattern. In Finnish, for example, the sequence LH— stressed light syllable followed by a heavy syllable — is generally avoided (Carlson 1978, Kiparsky 1991), but every polysyllable nevertheless begins with a bisyllabic foot, the strongest in the word, no matter what the composition of the first two syllables is. The constraint EDGEMOST is crucially involved; specifically, EDGEMOST(Hd-F; left;Wd), the member of the EDGEMOST family which requires the head foot of the prosodic word to be placed in initial position (see below §4.1. p.35 for details of formulation). EDGEMOST(Hd-F; left;Wd) evidently dominates whatever prominent or foot-shape constraint is responsible for the

---

<sup>18</sup> This situation can also be dealt with by aggressive treatment of the segmental input, as in ‘Trochaic Shortening’, whereby underlying HL is realized as LL (Prince 1990; Hayes 1991). Fijian provides a clear example (Dixon 1988; Hayes 1991), as does the phenomenon of trisyllabic shortening in English and other languages (Prince 1990). Here a PARSE-type constraint (PARSE- $\mu$ ) occupies a position inferior to FOOT-FORM in the hierarchy, so that a mora in the input will be left unparsed in order to get ~(LL)# out of /~HL/. Similar phenomena are discussed in §4.5 below and in Mester (in press).

<sup>19</sup> Poser’s analysis is configured so that coalescence is *blocked* when the second vowel of a sequence bears main-stress. Triggering and blocking are often interconvertible in this way; it all depends on what you think the ameliorative process is.

LH effect. Among the Yupik languages, it is typically the case that the quantity sensitivity of the basic pattern is restricted to the contrast V vs. VV; but in the first foot, arguably the prosodic head, CVC will count as heavy. (On this, see Hewitt 1992:§2 for extensive discussion, though from a somewhat different point of view; and Hayes 1991/1995, §6.3.8 for another perspective). Here EDGEMOST(Hd-F; left;Wd) dominates the constraint that limits moras to vocalicity; this is not unlike Tongan, in that the higher-level foot-placement constraint influences the basic interpretation of syllable structure. Similarly, it is often the case that the location of the main-stressed foot has a different directional sense from the rest of the foot pattern, a fact whose importance is emphasized in the work of Hulst 1984, 1992, which advocates top-down analysis of stress patterns in an operational framework. In Garawa, for example, the rhythmic pattern is built from bisyllabic trochees and every word begins with the main-stressed foot; but the rest of the word is organized with reference to the far end, right-to-left in iterative terms (Hayes 1980 working from Furby 1974). In Polish, also syllabic-trochaic, the principal foot lies at the end, but otherwise the pattern has a left-to-right orientation (Rubach & Booij 1985). Here again the constraint EDGEMOST(Head-F;rt;Wd), which governs main-stress distribution, dominates the constraint checking the directional uniformity of the pattern, call it SENSE(F), which is concerned with the lower hierarchical level of mere feet.<sup>20</sup> As we expect in the current context, reversing domination to SENSE(F) >> EDGEMOST(Head-F;rt;Wd) is entirely sensible, and defines the class of languages which is in accord with the spirit of Bottom-up Constructionism.

---

<sup>20</sup> Exact formulation of the directionality-sensitive principle is a matter of interest. One tack, noted by Green 1992 *inter alia* is to have the constraint sensitive to the location of *unfooted* material; thus a left-to-right operational iteration leaves the stray bits on the right side of the run; sometimes internally, in the case of bimoraic feet. It is also possible to deal directly with the matter and define a recursive checking procedure which is similar in spirit to the original iterative constructional conception (Howard 1973, Johnson 1972). Let there be a constraint EDGEMOST\* which is just like EDGEMOST except that it operates along these lines: string  $\alpha$  is more harmonic than string  $\beta$  under EDGEMOST\*(F,E,D) if  $\alpha$  is more harmonic than  $\beta$  under simple EDGEMOST(F,E,D); but if  $\alpha$  and  $\beta$  are not distinguished by simple EDGEMOST, because the edgemost foot of  $\alpha$ ,  $F_\alpha$ , is as well-located as the edgemost foot of  $\beta$ ,  $F_\beta$ , then resume the evaluation wrt EDGEMOST\* — nb. recursive reference — on  $\alpha \setminus F_\alpha$  and  $\beta \setminus F_\beta$ , where  $x \setminus y$  means ‘x with y removed’. This method of evaluation, which is like the Lexicographic Order on Composite Structures (LOCS) of Prince & Smolensky 1991, 1992, has been independently suggested by Itô, Mester, & McCarthy. Zoll 1992 offers yet another approach to directionality, in which an entire subhierarchy works across the word, in the spirit of Prince 1990. Further research is required to distinguish among these possibilities.

## 4. Generalization-Forms in Domination Hierarchies II

### Do Something Except When: Blocking, or The Theory of Profuseness

When specific conditions limit the scope of an otherwise broadly- applicable generalization, we have the second major form of constraint interaction. The operational imperative is to do something *except when* an adverse condition prevails or an adverse outcome would result. The default is to do; but there may be reasons not to do. In the literature, this is often understood as the *blocking* of a rule by a constraint; though the effect is also achieved when a later rule undoes part of what an earlier rule has accomplished, or when the Elsewhere Condition preserves the specific in the face of the general. This generalization structure emerges when a constraint that functions to encourage some easily identifiable state-of-affairs is dominated by another that can force a violation, ruling out the otherwise desired state-of-affairs in certain circumstances.

Clearly, this pattern of generalization is central to linguistic theory and appears in every branch of it. We will focus our discussion on basic elements of prosodic theory. Here are some examples of phenomena that earn the ‘do something except when’ interpretation:

#### (26) **Blocking of the Generally Applicable**

- a. A final syllable is extrametrical *except when* it is the only syllable in the word.
- b. A certain affix is a prefix *except when* the base is C-initial/V-initial (in which case, it is an infix).
- c. Stress is nonfinal *except when* the final syllable is the heaviest in the word.
- d. Rhythmic units are iambic *except when* there are only 2 syllables in the word.
- e. Rhythmic units are trochaic, and never of the form LH *except when* those are the only two syllables in the word.

Explicating such patterns via constraint domination will significantly advance the understanding of such fundamentalia as prosodic minimality, extrametricality, foot-typology, as well as the sometimes subtle interactions between them.

### 4.1 Edge-Oriented Infixation

Infixation comes in two varieties, each a mild variant of ordinary suffixation/prefixation, according to the typology of McCarthy & Prince 1990a. In both cases, the infix is a suffix/prefix which takes as its base not some morphological category (stem, root, etc.) but instead a phonologically-defined subdomain within the morphological category.

In one type, the affix attaches to a minimal word (foot-sized unit) that lies at the edge of the morphological base. For McCarthy & Prince 1990, the internal minimal word that forms the actual base of affixation is specified by positive Prosodic Circumscription. Thus, we find in the Nicaraguan language Ulwa that the personal possessive affixes are suffixed to the first foot, as in /kuhbil+ka/ → kuhkabil ‘his knife’ (Hale & Lacayo Blanco 1988). Reduplicative infixation may also follow this pattern. Samoan ‘partial reduplication’, for example, prefixes a syllable to the last foot, giving /σ+fa:gota/ → fa:go(góta) ‘fish<sub>v</sub> pl.’ (Broselow & McCarthy 1983, McCarthy & Prince 1990, Levelt 1991).

In the other type, the affix lies near an edge; its location can be determined by subtracting something (often a consonant or vowel) from the edge and prefixing or suffixing to the remainder. Characteristic instances are displayed below:

(27) **Tagalog Prefixal Infixation**

um+tawag	→ t- <i>um</i> -awag	‘call, pf., actor trigger’
um+abot	→ <i>um</i> -abot	‘reach for, pf., actor trigger’

(28) **Pangasinán Prefixal Reduplicative Infixation**

σ+amigo	→ a- <i>mi</i> -migo	‘friend/friends’
σ+libro	→ <i>li</i> -libro	‘book/books’

(29) **Chamorro Suffixal Reduplicative Infixation**

métgot+σ	→ métgo- <i>go</i> -t	‘strong/very strong’
buníta+σ	→ buníta- <i>ta</i>	‘pretty/very pretty’

In each of these examples, exactly one peripheral element is bypassed. For McCarthy & Prince 1990, Prosodic Circumscription identifies the unit that is subtracted from consideration, and the morphological operation applies to the residue of the form as though to an ordinary base. Ignoring a single peripheral element is extrametricality, and Prosodic Circumscription successfully formalizes extrametricality within a generalized theory of the prosodic conditioning of rules.

This approach is beset by a serious shortcoming: it cannot explain the relation between the *shape* of the affix and the manner of its placement. Infixes that go after initial C- are common throughout Austronesian, and they always have the form VC, never CV or V or CVC. Anderson (1972) observes that just when the affix is VC, infixation after a word-initial C results in a considerably more satisfactory syllable structure than prefixation would. The VC.CV entailed by simple prefixation (\**um.tawag*) is abandoned in favor of CV.CV. (*tu.mawag*). Of course, in the theory of Anderson’s day, such an observation — concerned with the relative markedness of different possible outputs — had to remain grammatically inert, and Anderson offered a 6-term transformational rule to describe the facts. Similarly, the observation plays no role in the Prosodic Circumscription account developed by McCarthy & Prince, which accordingly suffers from the same kind of formal arbitrariness diagnosed in the Dell-Elmedlaoui Algorithm. Optimality Theory, by contrast, allows us to make this insight the keystone of a significantly improved theory of edge-oriented infixation.

The relevant syllabic constraint is the one that discriminates against closed syllables: let us call it -COD and formulate it as follows:

(30) **-COD**

Syllables do not have codas.

Any closed syllable will violate this constraint, which we assume to be universally present in every grammar. It is often, though not always, in subordinate position in constraint rankings, dominated by faithfulness, so that a coda is forced when input like /CVC/ is faithfully parsed. (See §6 below.)



We need to generalize the notion *prefix* so that it can refer to items that are only approximately in first position in the string. The standard hard-line assumption is that a *prefix* always falls completely to the left of its base. Scalar evaluation allows us to soften the demand, using the now-familiar pattern of competition between alternatives that vary in degree of adherence to the ideal. Consider all possible affix placements: the nearer an affix lies to the left edge of the affix-base collocation, the more *prefixal* its status. The observed location of any particular affix will be that which best satisfies the entire set of constraints bearing on it, among them the requirement that it be a prefix or suffix.

The constraint needed to implement this idea is already at hand: it is EDGEMOST, familiar from the discussion of Tongan above, which we now formulate more precisely. Like the End Rule that it supplants, the predicate EDGEMOST takes several arguments. The relevant domain must be specified, as must the choice of edge (*Right*, *Left*), as must the item whose position is being evaluated. The schema for the constraint family, then, looks like this:

(31) **EDGEMOST** ( $\varphi$ ; E; D)

The item  $\varphi$  is situated at the edge E of domain D.

Crucially, violation of EDGEMOST is reckoned in designated units of measure from the edge E, where each intervening unit counts as a separate violation. In Tongan, the constraint comes out as EDGEMOST( $F'$ ;R; Wd), which asserts that the main foot ( $F'$ ) must be *Rightmost* in the Word. For conciseness, the Domain argument will be suppressed when its content is obvious. With this in hand, we can define the two affixal categories:

(32) Dfn. **Prefix**

A *prefix* is a morpheme  $\mathfrak{M}$  subject to the constraint EDGEMOST( $\mathfrak{M}$ ; L).


(33) Dfn. **Suffix**

A *suffix* is a morpheme  $\mathfrak{M}$  subject to the constraint EDGEMOST ( $\mathfrak{M}$ ; R).

Traditional prefixes or suffixes, which always appear strictly at their corresponding edges, are morphemes  $\mathfrak{M}$  for which EDGEMOST( $\mathfrak{M}$ ;L|R) dominates other conflicting constraints and thus always prevails. In the case of edge-oriented infixation, prosodic well-formedness constraints dominate EDGEMOST, forcing violations, which are minimal as always. For the Austronesian VC prefixes, the important interaction is  $-\text{COD} \gg \text{EDGEMOST}(af; L)$ . Consider the effects on the Tagalog prefix /um/, characterized by Schachter 1987 as signaling ‘Actor Trigger’ on the verb.

Let us examine first the case of a vowel-initial base, say /abot/ ‘reach for’. For purposes of this discussion, we assume that the candidate set produced by Gen will consist of all forms in which (1) the linear order of tautomorphemic segments is preserved, and (2) the segments of the affix are contiguous. We will also assume, following the general view of scholarship in the area, that there are vowel-initial syllables at the level of analysis we are concerned with, even though all surface syllables have onsets, glottal stop being the default filler. The following constraint tableau lays out the results.

## (34) Analyses of /um/ + /abot/

Candidates	-COD	EDGEMOST( <i>um</i> ; L)
a.  .u.ma.bot.	*	#∅
b. .a.um.bot.	* * !	#a
c. .a.bu.mot.	*	#ab !
d. .a.bo.umt.	*	#abo !
e. .a.bo.tum.	*	#abot !

The degree of violation of EDGEMOST is indicated informally by listing the string that separates *um* from the beginning of the word. The sign ‘#’ is included merely as a visual clue. The symbol ∅ refers to the empty string. The syllabically-illicit form *\*aboutmt* would of course be defeated on other grounds as well: it is included to emphasize the generality of the candidate set.

The form *umabot* is more harmonic than any competitor. Like all the best candidates, it has one -COD violation. (It is a fact of the language that this one violation cannot be circumvented by deletion or by epenthesis; this shows that -COD is dominated by both FILL and PARSE, as detailed in §6 below.) All ties on -COD are adjudicated by EDGEMOST. In the optimal form, the affix lies nearest to the beginning of the word: in this case, right *at* the beginning.

Consonant-initial bases provide the more interesting challenge. Consider the richest possible case, with an initial C-sequence, as in /um+gradwet/. (This datum is from French 1988.)

## (35) Analyses of /um + gradwet/


Candidates	-COD	EDGEMOST( <i>um</i> , L)
a. .um.grad.wet.	*** !	#∅
b. .gum.rad.wet.	*** !	#g
c.  .gru.mad.wet.	**	#gr
d. .gra.um.dwet.	**	#gra !
e. .gra.dum.wet.	**	#gra ! d
f. .grad.w...um...	**	#gra ! dw ...

Again, for completeness, we include *graumdwet* as a representative of the horde of syllabically disastrous candidates — including still more hopeless contenders such as *.umgra.dwet.* and *.u.m.gra.dwet.* — which are all rendered sub-optimal by constraints dominating -COD but of no relevance to the present discussion.

Here all the best competitors tie at two –COD violations each; these violations are due to intrinsic stem structure and cannot be evaded. Prefixing *um* to the whole base, as in (35.a), or after the initial C, as in (35.b), induces a third, fatal violation. Among the surviving forms, *grumadwet* wins on the grounds of superior affix position.

We have shown that certain prefixes are minimally infixes when prosodic constraints dominate the basic morphological principle of affix placement. McCarthy & Prince 1991, 1993, extending the present account, observe that the same explanation holds for edge-oriented *reduplicative* infixation. As an instance of suffixal  $\sigma$ , Chamorro *met.go-go-t* is prosodically superior to \**met.got.-got*. in that it more successfully avoids closed syllables; exactly the same reason that *t-u.m-a.wag* bests *um.-ta.wag*.

### (36) Suffixal Reduplication under Prosodic Domination

Candidates	–COD	EDGEMOST( $\sigma_{\text{aff}};R$ )
 .met.go.got.	**	t#
.met.got.got.	*** !	∅#

The widely-attested Pangasinán type *lilibro / amimigo*, in which the reduplicative prefix skips over an onsetless initial syllable, falls under the prosodic compulsion account as well. Here infixation — *a.mi.migo* vs. \**a.amigo* – avoids the V-V hiatus which straightforward prefixation would entail (cf. McCarthy & Prince 1986: 90). The active constraint is ONS, which discriminates against onsetless syllables.

The theory has a subtle and important range of consequences, whose existence was brought to our attention by John McCarthy: *certain types of infixation can only be reduplicative*. Under prosodic forcing, no phonemically-specified prefix can ever be placed after an initial onsetless syllable; the advantage of avoiding absolute initial position only accrues when reduplication is involved. To get a sense of this, note that under reduplication the failed candidate \$a\$a.migo, with pure prefixation, incurs two ONS violations, and the successful candidate \$a.mi.migo but one. (The ad hoc siglum \$ marks a missing Onset.) Reduplication has the unique pathology of copying an onset violation. By contrast, a fixed-content affix like (fictive) /ta/ leads to just one ONS violation wherever it is placed: pseudo-Pangasinán *ta\$a.migo* is not ONS-distinct from \$a.ta.migo. Since ONS is indifferent, EDGEMOST will demand full left-placement. Since all the known cases of post-onsetless syllable infixation are in fact reduplicative, we have a striking argument in favor of the present approach. The argument applies in a kind of mirror-image or dual form to suffixal infixation as well. The Chamorro *metgot* type can only be reduplicative. Consider, for example, the effects of placing an imaginary affix /ta/ in and around a form like *metgot*: *met.got.ta.*, *metgo.tat*. All such candidates agree on the extent of –COD violation, as the reader may verify; therefore EDGEMOST(*ta*;R) compels

exterior suffixation.<sup>21</sup> To firmly establish the claim that fixed-content morphemes cannot be forced into infixation after initial onsetless syllables or before final C, more must be done: one needs to run through all relevant affix patterns, checking the effects of contact between the edges of the affix with the base. In addition, the reduplicative pattern must be nailed down. Detailed analysis is undertaken in McCarthy & Prince 1993: §7, to which the reader is referred.

We have, then, the beginnings of a substantive theory of infixability, a theory in which prosodic shape modulates the placement of morphemes. Edge-oriented infixation arises from the interaction of prosodic and morphological constraints. The principal effect — interior placement — comes about because EDGEMOSTness is a gradient property, not an absolute one, and violations can be forced. It follows from the principles of the harmonic ordering of forms that violations in the output are minimal. Consequently, such infixes fall *near* the edge, as near as possible given the dominant constraints. Edge-oriented infixation can be construed in the ‘Do-Something-Except-When’ style of descriptive language, should that prove illuminating: the affix falls at the edge *except when* a prosodic constraint can be better met inside. The theory, of course, recognizes no distinction between ‘except when’ and ‘only when’ — blocking and triggering — but deals only in the single notion of constraint domination.

The internalizing effects attributed to extrametricality follow, on this view, from constraint interaction and from the way that constraints are defined. There is no formal mechanism called Extrametricality or (negative) Prosodic Circumscription to which the analysis appeals. This suggests the general hypothesis, natural within the context of Optimality Theory, that what we call Extrametricality is no more than the name for a family of effects in which Edgemostness interacts with other prosodic constraints. We pursue this line in the following two sections as we explore more instances of the *except when* configuration, showing that key properties of extrametricality, thought to be axiomatic, follow from this re-conception.

## 4.2 Interaction of Weight Effects with Extrametricality

Certain varieties of Hindi show an interaction between weight and nonfinal placement of stress which sheds further light on the interaction of gradient edgmostness and other factors operative in prosodic patterning. First, we provide some background on “unbounded” “stress” systems; then we turn to the revelatory twists of Hindi prosody.

### 4.2.1 Background: Prominence-Driven Stress Systems

Stress systems typically reckon main-stress from a domain edge, often enhancing an edgmost or near-edgmost syllable or foot. There are also stress systems that call on EDGEMOST but make no use of binary structure to define the position of main word-stress: instead the additional determining

---

<sup>21</sup> Observe that in all the cases discussed we can assume that the entire package of syllabic constraints, including both ONS and –COD, dominates the morphological conditions on affixation; one or the other member of the package turns out to be relevant depending on what the content of the affix is and whether it is a prefix or stem. This idea figures centrally in McCarthy & Prince 1993, where the Optimality theoretic scheme “prosody dominates morphology” is proposed as the account of what makes morphology prosodic.

factor is *syllable weight*. In the canonical cases, main stress falls on the leftmost/rightmost heavy syllable (pick one); otherwise, lacking heavies in the word, on the leftmost/rightmost syllable (pick one). Systems like these have been called “unbounded” because the distance between the edge and the main-stress knows no principled limits and because metrical analysis has occasionally reified these unbounded spans as feet (Prince 1976, 1980; Halle & Vergnaud 1987; Hayes 1980, 1991/1995). The best current understanding, however, is that what’s involved is not a foot of unbounded magnitude (presumed nonexistent), but a kind of prominent enhancement that calls directly on contrasts in the intrinsic prominence of syllables. These then are prominence-driven systems, in which a word’s binary rhythmic structure is decoupled from the location of main word-stress. (For discussion, see Prince 1983, 1990; Hayes 1991/1995.)

Two basic constraints are involved. First, it is necessary to establish the relation between the intrinsic prominence of syllables and the kind of elevated prominence known as stress. There are a number of ideas in the literature as to how this is to be done (Prince 1983, 1990, McCarthy & Prince 1986, Davis 1988ab, Everett 1988, Zec 1988, Goldsmith & Larson 1990, Hayes 1991/1995, Goldsmith 1992, Larson 1992), none perhaps entirely satisfactory. Generalizing over particular representational assumptions, we can write, following essentially McCarthy & Prince 1986:9,

(37) **Peak-Prominence** (PK-PROM)

Peak( $x$ ) > Peak( $y$ ) if  $|x| > |y|$ .

By PK-PROM, the element  $x$  is a better peak than  $y$  if the intrinsic prominence of  $x$  is greater than that of  $y$ . This is the same as the nuclear-Harmony constraint HNUC formulated above, which holds that higher sonority elements make better syllable peaks.

The second relevant constraint determines the favored position of the prominence-peak or main stress of the word. It is nothing other than the familiar EDGEMOSTness.

(38) **EDGEMOST**(pk; L|R; Word)

A peak of prominence lies at the L|R edge of the Word.

We use ‘Word’ loosely to refer to any stress domain; as before, EDGEMOST is subject to gradient violation, determined by the distance of the designated item from the designated edge.

To see how these constraints play out, let us consider a simple prominence-driven system such as “stress the rightmost heavy syllable, else the rightmost syllable.” Here we have

(39) PK-PROM  $\gg$  EDGEMOST(pk;R)

If there are no heavy syllables in the word, the rightmost syllable faces no competition and gains the peak. The results are portrayed in the following tableau:

(40) **Right-Oriented Prominence System**. No Heavy Syllables:

Candidates	PK-PROM	EDGEMOST(pk;R)
☞ L L L <b>́</b>		∅#
L L <b>́</b> L		σ# !
L <b>́</b> L L		σσ# !
<b>́</b> L L L		σσσ# !

Here PK-PROM plays no role in the decision, since all candidates fare equally on the constraint. This kind of data provides no argument for ranking the constraints; either ranking will do. With heavy syllables in the string, the force of constraint PK-PROM becomes evident:

(41) **Right-Oriented Prominence System**, with heavy syllables.

Candidates	PK-PROM	EDGEMOST(pk;R)
L H H <b>́</b>	<b>́</b>	
☞ L H <b>́</b> L	́	σ#
L <b>́</b> H H L	́	σσ# !
<b>́</b> H H L	<b>́</b> !	σσσ#

With the other domination order, a strictly final stress location would always win. With PK-PROM dominant, candidates in which a heavy syllable is peak-stressed will eclipse all those where a light syllable is the peak. When several potential peaks are equivalent in weight, or in intrinsic prominence construed more generally, the decision is passed to EDGEMOST, and the surviving candidate containing the peak nearest the relevant edge is evaluated as optimal; exactly the generalization at hand.

### 4.2.2 The Interaction of Weight and Extrametricality: Kelkar's Hindi

Certain dialects of Hindi/Urdu display an interesting variant of the prominence-driven pattern of edgemostness.<sup>22</sup> From the work of Kelkar (1968), Hayes (1991/1995:276-278) has constructed the following generalization:

(42) **Kelkar's Hindi**

“Stress falls on the heaviest available syllable, and in the event of a tie, the rightmost nonfinal candidate wins.”

(Hayes 1991/1995:276)

The first complication is that this variety of Hindi (or Urdu) recognizes three degrees of syllable weight or intrinsic prominence; hence Hayes's ‘heaviest’ holding the place of the usual ‘heavy’. The ordering of weight-classes is as follows:

(43) **Heaviness Scale**      |CVVC,CVCC| > |CVV,CVC| > |CV|

Hayes suggests that the superheavy syllables are trimoraic, yielding the scale | $\mu\mu\mu$ | > | $\mu\mu$ | > | $\mu$ |. Whatever the proper interpretation may be, the heaviness scale fits directly into the constraint PK-PROM.

The effects of PK-PROM may be seen directly in forms which contain one syllable that is heavier than all others:

(44) **Heaviest wins**

a. .ki.d <sup>h</sup> ár.	. $\mu\mu$ .	>	. $\mu$ .	‘which way’
b. .ja.náab.	. $\mu\mu\mu$ .	>	. $\mu$ .	‘sir’
c. .as.báab.	. $\mu\mu\mu$ .	>	. $\mu\mu$ .	‘goods’
d. .ru.pi.áa.	. $\mu\mu$ .	>	. $\mu$ .	‘rupee’
e. .réez.ga.rii.	. $\mu\mu\mu$ .	>	. $\mu\mu$ .	‘small change’

(All examples here and below are from Hayes 1991/1995.)

The second complication in the Hindi pattern is the avoidance of stress on final syllables. This is a very commonly encountered phenomenon in stress systems of all kinds, typically attributed to various forms of extrametricality, stress-shift, and de-stressing. We formulate the basic constraint as NONFINALITY as follows:

---

<sup>22</sup> Judgments of stress in Hindi are notoriously delicate and unstable, a consequence of dialectal variation and the non-obviousness of whatever events and contrasts the term ‘stress’ actually refers to in the language. Therefore it is essential to distinguish the observations of distinct individuals and to seek non-impressionistic support for the claims involved. Hayes (1991: 133-137, 236-237) provides careful analysis along these lines.

(45) **NONFINALITY**

The prosodic head of the word does not fall on the word-final syllable.

By ‘prosodic head’ we mean the prosodically most prominent element, here the main stress. NONFINALITY is quite different in character from extrametricality; it focuses on the well-formedness of the stress peak, not on the parsability of the final syllable.<sup>23</sup> Furthermore, it is a substantive stress-specific constraint, not a general mechanism for achieving descriptive ‘invisibility’ (Poser 1986).

When heaviness alone does not decide between candidates, the position of the peak is determined by the relation NONFINALITY  $\gg$  EDGEMOST. It is more important for the peak to be nonfinal than for it to be maximally near the edge. Exactly as in the simple prominence-driven systems, however, the package of positional constraints is completely dominated by the weight-measuring PK-PROM. Here are some examples illustrating the positional effects (syllables of the heaviest weight class in a word are in roman type):

(46) **Positional Adjudication among Equals**

a.	$\mu$	.sa. <b>mí</b> .ti.	‘committee’
b.	$\mu\mu$	.ru. <b>káa</b> .yaa. <b>pús</b> .ta.kee. .roo. <b>záa</b> .naa.	‘stopped (trans.)’ ‘books’ ‘daily’
c.	$\mu\mu\mu$	. <b>áas</b> .mãã.jaah. .aas. <b>máan</b> .jaah.	‘highly placed’ ‘highly placed (var.)’

The full constraint hierarchy runs PK-PROM  $\gg$  NONFINALITY  $\gg$  EDGEMOST. The following tableaux show how evaluation proceeds over some typical examples.

---

<sup>23</sup> NONFINALITY does not even imply by itself that the literally last syllable is unstressed. Representational nonfinality can be achieved in the manner of Kiparsky 1992 by positing an empty metrical node or grid-position (analogous to the ‘silent demi-beat’ of Selkirk 1984) after the final syllable within the stress domain. Use of empty structure is proposed in Giegerich 1985 and Burzio 1987 for various purposes, and is explored in Kiparsky 1992 under the name of ‘catalexis’, in connection with preserving Foot Binarity (q.v.inf.). Here we want empty metrical positions to be unavailable; clearly, they are proscribed by a constraint of the FILL family, and we will tacitly assume that this constraint is undominated in the grammars under discussion.



(47) **Light vs. Light:** /samiti/

Candidates	PK-PROM	Position	
		NONFINALITY	EDGEMOST
.sa.mi.tí.	.ú.	* !	∅#
☞ .sa.mí.ti.	.ú.		σ#
.sá.mi.ti.	.ú.		σσ# !

The form .sa.mí.ti. is optimal because it has a nonfinal peak that is nearest the end of the word.

(48) **Heavy vs. Light:** /kid<sup>h</sup>ar/

Candidates	PK-PROM	Position	
		NONFINALITY	EDGEMOST
☞ .ki.d <sup>h</sup> ár.	.úμ.	*	∅#
.kí.d <sup>h</sup> ar.	.ú. !		σ#

The optimal form .ki.d<sup>h</sup>ár. violates NONFINALITY, but it wins on PK-PROM, which is superordinate.

(49) **Heavy vs. Heavy vs. Light:** /pustakee/

Candidates	PK-PROM	Position	
		NONFINALITY	EDGEMOST
.pus.ta.kée.	.úμ.	* !	∅#
.pus.tá.kee.	.ú. !		σ#
☞ .pús.ta.kee.	.úμ.		σσ#

The form .pús.ta.kee. is the worst violator of EDGEMOSTness among the candidates, but it bests each rival on a higher-ranked constraint.

(50) **Contest of the Superheavies:** /aasmããjaah/

Candidates	PK-PROM	Position	
		NONFINALITY	EDGEMOST
aas.mãã.jáah	μμμ	* !	∅#
aas.mãã.jaah	μμ !		σ#
☞ áas.mãã.jaah	μμμ		σσ#

Here again, the optimal candidate is the worst violator of edgemostrness, but its status is assured by success in the more important confrontations over weight and nonfinality.

The stress pattern of Kelkar's Hindi shows that extrametricality can be 'canceled' when it interferes with other prosodic constraints. It is rarely if ever the case that final syllables are categorically extrametrical in a language; rather, prominence is nonfinal *except when* being so entails fatal violation of higher-ranked constraints. This behavior is exactly what we expect under Optimality Theory. In the familiar view, of course, such behavior is a total mystery and the source of numerous condundra, to be resolved by special stipulation; for if extrametricality is truly a rule assigning a certain feature, there can be no explanation for why it fails to apply when its structural description is met.

### 4.3 Nonfinality and Nonexhaustiveness

The exclusion of *word-final syllables* from prosodic structure is the prototypical extrametricality effect. Latin provides the touchstone example, and parallels can be multiplied easily.<sup>24</sup> Writing the extrametricality rule to apply word-finally leads immediately to the basic quirk of the theory: *monosyllabic* content words receive stress without apparent difficulty. Since the unique syllable of the monosyllable is indubitably final, it should by all rights be extrametrical. Why is this syllable different from all others? The following examples illustrate the situation, where ⟨...⟩ encloses the extrametrical material:

#### (51) Extrametricality in Latin

- a. cór⟨pus⟩            \*corpús  
 b. méns                \*⟨mens⟩

<sup>24</sup> In words of length two syllables or longer, Latin places main word-stress on the penult if it is heavy or if it is the first syllable in the word, otherwise on the antepenult. This array of facts is standardly interpreted to mean that final syllables are completely extrametrical — outside foot structure. Bimoraic trochees are applied from left to right on the residue of extrametricality; the last foot is the strongest (Hayes 1980, 1987).

This state of affairs arises from an interaction exactly parallel to the one that ‘revokes extrametricality’ in Hindi. NONFINALITY is simply not the *ne plus ultra* of the system; it can be violated.

The dominant, violation-forcing constraint is not far to seek. Relations must be established between the categories of morphology and those of phonology. These take the form of requirements that any member of a certain morphological category (root, stem, word) must be, or correspond to, a phonological category, typically the prosodic word *PrWd*. (See Liberman & Prince 1977; Prince 1983; McCarthy & Prince 1986, 1990, 1991ab, 1993; Nespor & Vogel 1986; Inkelas 1989.) The *PrWd* is composed of feet and syllables; it is the domain in which “main stress” is defined, since every *PrWd* contains precisely one syllable bearing main stress. As in many languages, Latin requires that the lexical word be a prosodic word as well. Following McCarthy & Prince 1991ab, we can put the morphology/phonology interface constraint like this, with one parameter:

(52) **LX $\approx$ PR** (*MCat*)

A member of the morphological category *MCat* corresponds to a *PrWd*.

Another line of approach is to demand that the left or right *edge* of a morphological category match to the corresponding edge of the relevant phonological category (Selkirk 1986, Chen 1987, McCarthy & Prince 1993). For present purposes it is not necessary to pursue such refinements of formulation, although we return in §7, p. 114ff., to the virtues of edge-reference.

All words of Latin satisfy LX $\approx$ PR; not all final syllables are stressless. (Indeed, on the standard view of Latin prosodic structure, a final syllable is included in stress structure only in monosyllables.) NONFINALITY is violated exactly when LX $\approx$ PR is at stake. We deduce that LX $\approx$ PR  $\gg$  NONFINALITY. It remains to formulate a satisfactory version of the constraint from the NONFINALITY family that is visibly active in Latin. For present purposes, the following will suffice:

(53) **NONFINALITY**

The head *foot* of the *PrWd* must not be final.

This is related to NONFINALITY (45) §4.2.2, p.42, which deals with peaks of stress — syllabic heads of *PrWd* — but not identical with it. We will bring them together shortly, in order to deal with the subtler interactions between nonfinality and foot-form restrictions.

The effect of the constraint hierarchy on monosyllables is illustrated in this tableau:

(54) **The Parsed Monosyllable of Latin**

Candidates	LX $\approx$ PR	NONFINALITY
☞ [ (méns) <sub>F</sub> ] <sub>PrWd</sub>		*
⟨mens⟩	* !	

The constraint  $LX \approx PR$  word thus ‘revokes extrametricality’ when content-word monosyllables are involved.

It is instructive to compare the present approach with the standard conception, due to Hayes, which holds that extrametricality is a feature assigned by rule as part of the bottom-up process of building prosodic structure. Under Bottom-up Constructionism, there must be a strict serial order of operations:

1. Extrametricality marking must take place: this prepares the syllabified but footless input for further processing.
2. Feet are then formed, determining the location of stressed and unstressed syllables.
3. Higher Order structure is then built on the feet — i.e., the Prosodic Word is formed — and the location of main stress is determined.

Under this plan of action, it is essential that extrametricality be assigned *correctly* at the very first step. If monosyllabic input is rendered entirely extrametrical at step #1, then Prosodic Word Formation (step # 3) will have no feet to work with, and will fail. To avoid this disastrous outcome, a caveat must be attached to the theory of extrametricality to ensure that the fatal misstep is never taken. Hayes formulates the condition in this way:

**(55) Nonexhaustivity**

“An extrametricality rule is blocked if it would render the entire domain of the stress rules extrametrical.” (Hayes 1991/1995: 58)

It is an unavoidable consequence of Bottom-up Constructionism that condition (55) must be stated as an independent axiom of theory, unrelated to any other constraints that bear on prosodic wellformedness. Its existence is entirely due to the theory’s inability to recognize that  $LX \approx PR$  is a constraint on the *output* of the system, a condition that must be met, and not the result of scanning input for suitable configurations and performing Structural Changes on them. The putative rule assigning PrWd status cannot be allowed to fail due to lack of appropriate input.<sup>25</sup> The Axiom of Nonexhaustivity is not motivated by restrictiveness or any other such higher explanatory motive. Its motivation is strictly empirical; remove it and you have an equally restrictive theory, but one which predicts the opposite treatment of monosyllables.

Nonexhaustiveness appears not to be part of any *general* theory of extrametricality. Hewitt & Prince (1989), for example, argue that extrametricality with respect to tonal association may indeed exclude entire monosyllabic domains. Hayes is careful to refer to the notion “*stress domain*” in his statement of the condition. The proposal offered here makes sense of this: the integrity of the stress domain is guaranteed by the theory of the morphology/phonology interface, as encoded in

---

<sup>25</sup> Compare, in this regard, the discussion of relative clause formation in Chomsky 1965, where it is noted that it is insufficient to say that the rule of relative clause formation is *obligatory*, because nothing guarantees that it will be able to apply at all (\*the man that the house looks nice). Compare also the notion of “positive absolute exception” in Lakoff 1965, a rule whose structural description *must be met*. These phenomena are diagnostic of deep failure in the simple re-write rule conception of grammar, since remedied. In the case of relative clauses, it is clear that the syntax is entirely free to create structures in which no wh-movement can apply, because independent principles of interpretation, defined over the *output* of the syntax, will fail in all such forms, ruling them out.

LX≈PR. When that particular theory is not involved, as in certain tonal associations, there is no reason to expect nonexhaustiveness, and we do find it. Similarly, the end-of-the-word bias of stress-pattern extrametricality is not mirrored in other phenomena which ought to fall under the theory of extrametricality (were there to be one). For example, tonal extrametricality (Prince 1983, Pulleyblank 1983) is not restricted to final position; nor is edge-oriented infixation (McCarthy & Prince 1986, 1990). We expect this: extrametricality is not a unified entity, but rather a diverse family of consequences of the gradience of EDGEMOSTNESS. In the subtheory pertaining to stress, NONFINALITY is the principal, perhaps only, constraint interacting with EDGEMOSTNESS. In other phenomenal domains besides stress, other constraints are at play, shown above in the case of edge-oriented infixation, §4.1.

Nonexhaustiveness, then, emerges from constraint interaction. What of the other properties that have been ascribed to formal extrametricality? There are four, and in each case, we would argue, what is correct about them follows from the constraint interaction analysis. Let's take them in turn.

(56) Property 1: **Constituency**

“Only constituents (e.g. segment, mora, syllable, foot, phonological word) may be marked as extrametrical.” (Hayes 1991/1995: 57).

We suggest that this property has nothing to do with extrametricality *per se* but rather with the substantive constraint that pushes the relevant item off an edge. Constraints on stress, for example, deal in syllables quite independently of extrametricality. When the relevant constraint is from a different domain, it may well be that constituency is irrelevant; in edge-oriented infixation, for example, as analyzed above in §4.1, the constraint –COD can force prefixes away from the initial edge of the word, over consonant sequences that needn't be interpreted as unitary constituents (“onsets”).

(57) Property 2: **Peripherality**

“A constituent may be extrametrical only if it is at a designated edge (left or right) of its domain.” (Hayes, *ibid.*).

This is because the phenomena gathered under the name of extrametricality have to do with items that are positioned by the constraint EDGEMOST — prominences, feet, tones, affixes. If by extrametrical, we mean “unparsed into the relevant structure”, then there are many other situations where constraints force nonparsing. Hayes's “weak local parsing”, for example, compels unparsed syllables to separate binary feet (the similarity to extrametricality is recognized in Hammond 1992). Syllables may be left unparsed internally as well as peripherally because of restrictions on the quantitative shape of feet (the “prosodic trapping” of Mester 1992). Similar observations may be made about segmental parsing. Many kinds of constraints can lead to nonparsing; we assert that there is no reason to collect together a subset of them under the name of extrametricality.

(58) Property 3: **Edge Markedness**

“The unmarked edge for extrametricality is the right edge.” (Hayes, *ibid.*)

As noted, this is true only for stress, not for tone or affixation. The explanation must lie in the properties of stress, not in a theory of the treatment of edges.

(59) Property 4: *Uniqueness*.

Only *one* constituent of any type may be extrametrical.

This is a classic case of constraint interaction as we treat it. Extrametricality arises, for example, when NONFINALITY  $\gg$  EDGEMOSTNESS. It follows that EDGEMOSTNESS is violated when extrametrical material is present. Because of the way Harmony is evaluated (HOF: §5), such violations must be minimal. Under NONFINALITY, this will commonly mean that only one element is skipped over or left unparsed, the minimal violation of EDGEMOSTNESS. Thus, in many cases — enough to inspire belief that a parochial principle is involved — the unparsed sequence will be a constituent.

We conclude that there are strong reasons to believe that extrametricality should be retired as a formal device. Since its basic properties submit to explanation in the substantive domain under scrutiny here, it is worthwhile to pursue the argument into the other areas where it has proved to be such a useful tool of analysis. (For further exploration of nonfinality and related edge effects, see Hung 1993.)

Demoting nonexhaustivity from clause-of-UG to epiphenomenon of interaction raises an important issue, however: the universality (or at least generality) of the effect is not directly accounted for. What ensures that  $LX \approx PR$  outranks NONFINALITY? Why not have it the other way around in the next language over? It may be that some further condition is required, restricting the place of  $LX \approx PR$  in constraint hierarchies. Have we therefore exchanged one stipulation for another, failing to net an overall gain at the bottom line?

Reviewing the spectrum of possible responses, it's clear that we're not in a simple tit-for-tat situation. Suppose we straightforwardly call on some principle relevant only to  $LX \approx PR$ ; for example, that it must sit at the top of all hierarchies, undominated. Any such condition fixes the range of relations between  $LX \approx PR$  and many other constraints, not just NONFINALITY, so the cost of the stipulation is amortized over a broad range of consequences having to do with the prosodic status of lexical items. Thus, even with the most direct response, we put ourselves in a better position than the adherent of pure axiomatic status for nonexhaustiveness as a property local to extrametricality.<sup>26</sup>

---

<sup>26</sup> More optimistically, we can expect to find principles of universal ranking that deal with whole *classes* of constraints; in which desirable case, the nonexhaustiveness effect would fully follow from independent considerations. Also worth considering is the idea that there are no principles involved at all and the predominance of the cited ranking is due to functional factors extrinsic to grammar, e.g. the utility of short words. (This comports as well with the fact that conditions on word minimality can differ in detail from language to language, including or excluding various categories from the 'lexical word': see the discussion of Latin which immediately follows; implying a family of related constraints, rather than a single one.) Grammar allows the ranking to be easily learnable from the abundant data that justifies it. On this view, nothing says that NONFINALITY must be dominated; but it is easy to observe that it is; and there are extragrammatical, functional reasons why it is useful for it to be. Note too that in setting the rank of  $LX \approx PR$ , we construct the needed restriction from UG-building tools already needed; as opposed to tacking nonexhaustiveness on as a *sui generis* codicil to some mechanism.

It is useful to compare the kind of ranking argument just given with a familiar form of argument for rule ordering in operational theories. One often notes that if Rule A and Rule B were to apply simultaneously, then the conditions of Rule A must be written into Rule B; whereas if Rule A strictly precedes Rule B, the two can be disentangled, allowing the development of an appropriately restrictive theory of type A and type B rules.<sup>27</sup>

Here we have seen that the empirical generalization about extrametricality (that it holds *except when* it would obliterate a monosyllable) emerges properly from the ranking of independent constraints. The rule-ordering theory of extrametricality, by contrast, exhibits a pathological quirk similar to the one that affects simple non-ordering operational theories. Information proper to one rule must be written into another, solely to get the right outcome: the Nonexhaustivity axiom embodies a covert reference to PrWd-formation. A grammar of re-write rules is simply not suited to the situation where a rule *must apply*, where its structural description *must be met* by all inputs so that all outputs conform to its structural change (see fn. 25 above). Inserting special conditions into rules so that this happens to happen is no answer. Rule-ordering must therefore be abandoned in favor of constraint ranking, for the same reason that simultaneity was previously abandoned in favor of rule ordering.

### 4.3.1 Nonfinality and the Laws of Foot Form: Raw Minimality

Latin displays a typical minimality effect of the type made familiar by work in Prosodic Morphology: the language lacks monomoraic words. Here is a list of typical monosyllabic forms (Mester 1992:19-20):

#### (60) Latin Monosyllables

Category	Exempla	Glosses
a. N	<i>mens, cor, mel, rē, spē, vī</i>	‘mind, heart, honey: nom.; thing, hope, force: abl’
b. V	<i>dō, stā, sum, stat</i>	‘I give, stand, am; he stands’
c. Pron.	<i>mē, sē, tū, is, id, quis</i>	‘1sg.-acc., 3-refl.-acc.; you-sg-nom; he, it, who-nom.’
d. Conj.	<i>nē, sī, cum, sed</i>	‘lest, if, when, but also’
e. P	<i>ā, ē, prō, sub, in, ab</i>	‘from, out of, in front of, under, in, from’

We have then *cum, mens, re* and *rem* (acc.), all bimoraic, but no *\*rē* (Mester 1992, citing Kuryłowicz 1968, Allen 1973:51). A morpheme like *-quē* ‘and’ can only be enclitic. The standard account points to the prosody-morphology interface constraint  $LX \approx PR$  as the source of the restriction (Prince 1980; McCarthy & Prince 1986 *et seq.*). The PrWd must contain at least one foot; a foot will contain at least two moras; hence, lexical words are minimally bimoraic. The deduction rests on the

---

<sup>27</sup> Notice that the argument has nothing to do with ‘redundancy’, which (here as elsewhere) is nothing more than a diagnostic indication that greater independence could be achieved; and with that, explanation.

principle *Foot Binariness*, which we state in (61), following the formulation of McCarthy & Prince 1986.<sup>28</sup>

(61) **Foot Binariness (FTBIN)**

Feet are binary at some level of analysis ( $\mu$ ,  $\sigma$ ).

Foot Binariness is not itself a direct restriction on minimal foot size; it defines a general property of structure. (Indeed the obvious virtue of the prosodic theory of minimality is that it obviates the need for such a thing as a ‘minimal word constraint’; but see Itô & Mester 1992.) Because syllables *contain* moras, Foot Binariness entails that the smallest foot is bimoraic.

One then argues that the lexicon of Latin cannot contain *e.g.* a noun /re/ because any such item wouldn’t be assigned foot-structure (assuming FTBIN), and therefore wouldn’t be assigned PrWd status. Since (on this view, not ours) a version of the constraint  $LX \approx PR$  holds categorically of the output of the lexicon, rejecting all violators, it happens that a potential lexical form like /re/ would underly no well-formed output and is therefore impossible, or at least pointless as a lexical entry. On this view, the constraint  $LX \approx PR$  assigns absolute ill-formedness to the output; in consequence, an input form /re/ yields no output at all.

In the present theory, failure to satisfy a constraint does not entail ill-formedness, and every input is always paired with an output: whichever analysis best satisfies the constraint hierarchy under HOF. We must therefore delve deeper to see how *absolute* ill-formedness — lack of an effective output — could emerge from a system of interacting constraints which always selects at least one analysis as optimal; this (or these) being by definition the output of the system.

The key, we suggest, is that among the analyses to be evaluated is one which assigns no structure at all to the input: the *Null Parse*, identical to the input. The Null Parse will certainly be superior to some other possibilities, because it vacuously satisfies any constraint pertaining to structures that it lacks. For example, FTBIN says *if* there is a foot in the representation, *then* it must be binary; violations are incurred by the *presence* by nonbinary feet. The Null Parse therefore satisfies FTBIN, since it contains no feet of any kind. Similar remarks hold for syllable structure constraints such as ONS, because the Null Parse contains no syllables; for structural constraints such as FILL, which demands that empty nodes be absent (they are). Of course, the Null Parse grossly fails such constraints as PARSE, which demands that segments be prosodically licensed, to use Itô’s term, because the input will always contain segments. The Null Parse will fail  $LX \approx PR$  when the input

---


<sup>28</sup> Foot Binariness is generalized from the account of Estonian foot structure in Prince 1980. The foot of Estonian is there defined as  $F = uu$ , where  $u$  is a variable ranging uniformly over  $\mu$  and  $\sigma$ . It is shown that this definition holds of the stress system, and the entailed minimality result — all monosyllables are bimoraic, hence overlong — is shown to hold as well. With such a foot, there are parsing ambiguities in certain circumstances —  $H\sigma$  may be  $(H)\sigma$  or  $(H\sigma)$  — and it is proposed that the maximal (bisyllabic) analysis is the one taken. The Estonian type of foot has been dubbed the “Generalized Trochee” in Hayes 1991, and new applications of it are reported; cf. Kager 1992abc for further exploration. Since iambic feet also meet the description ‘bimoraic or bisyllabic’, we go beyond the realm of the trochaic and, with McCarthy & Prince 1986 and Prince 1990, demand that FTBIN hold of all feet, regardless of headedness. Of course, FTBIN excludes the ‘unbounded’ feet of early metrical theory.



string *is* a lexical category, because the constraint applies to all items in the category; it says that all such items must be parsed as prosodic words.

A direct phonological assault on the Latin type of minimality would run like this: suppose that FTBIN  $\gg$  LX $\approx$ PR, with PARSE the lowest-ranked relevant faithfulness constraint. We have then the following:

### (62) Optimality of the Null Parse

Candidates	FTBIN	LX $\approx$ PR	...	PARSE
 re	<i>vacuous</i>	*	...	*
[ ( ré ) <sub>F</sub> ] <sub>PrWd</sub>	* !		...	

In the language of actions and exceptions, it would be said that a lexical word becomes a prosodic word *except when* it is monomoraic; the “assignment” of PrWd status to subminimals is *blocked* by FTBIN.<sup>29</sup>

Clearly, then, there exist constraint hierarchies under which the Null Parse is optimal for certain inputs. The Null Parse, however, is uniquely unsuited to life in the outside world. In the phonological realm alone, the principle that unlicensed material is phonetically unrealized (our take on the “Stray Erasure” of McCarthy 1979, Steriade 1982, Itô 1986, 1989) entails that any item receiving the Null Parse will be entirely silent. In a broad range of circumstances, this would render an item useless, and therefore provide the basis for an explanation for its absence from the lexicon. (Observe that some similar account is required under any theory, since nothing in the formal nature of a lexical entry prevents its from being entirely empty of phonetic content in the first place.)

A more bracing line of attack on the general problem is disclosed if we broaden our attention to include morphology. As with phonology, morphological structure can be understood as something that the input lacks and the output has, the product of parsing. In the simple case of affixation, a structure like [Root Af]<sub>Stem</sub> is an output possibility, sometimes the best one, related to a hierarchically unstructured input {*root<sub>i</sub>*, *affix<sub>j</sub>*} which merely collects together the constituent morphemes. Here too, though, the Null Parse must be among the options, and will be superior to some others, often to *all* others. For example, if *affix<sub>j</sub>* is restricted to combining with a certain subset of the vocabulary, and *root<sub>i</sub>* is not in this set, then the **non**combination of the morphological Null Parse  $\langle root_i, affix_j \rangle$  is going to be superior to the parsed *mis*combination [root<sub>i</sub> affix<sub>j</sub>]<sub>Stem</sub>. A real-life example is provided by English comparative-superlative morphology, which attaches only to (one-foot) Minimal Words:

<sup>29</sup> One looseness of formulation here is the exact category to which ‘LX’ in the constraint refers. John McCarthy reminds us that there are no alternations of the form  $\emptyset$  (nom.)  $\sim$  *rem* (acc.) from /re+m/, even though the inflected form is legitimately bimoraic. One could argue that the absence of such alternations is not properly grammatical, but functional in origin, considering the serious communicative problems entailed by lack of a nominative form. One might also argue that this shows that the Latin constraint must actually apply to the category (noun) *root*. There are however a number of further issues concerning how minimality plays out over the various morphological categories; see Mester 1992, especially §2.2, for analysis.


thus [violet-er], parsed, is inferior to ⟨violet, er⟩, which evades the one-foot constraint by evading attachment. Similarly, the input {write, ation} is best left untouched, but the analysis of {cite, ation} is optimally a parsable word. (For discussion of related cases, see McCarthy & Prince 1990, 1993).

Failure to achieve morphological parsing is fatal. An unparsed item has no morphological category, and cannot be interpreted, either semantically or in terms of higher morphological structure. This parallels the phonetic uninterpretability of unparsed segmental material. The requirements of higher order prosody will parallel those of higher order morphology and syntax: a phonological Null Parse, which assigns no Prosodic Word node, renders a word unusable as an element in a Phonological Phrase (Selkirk 1980, Nespor & Vogel 1986, Inkelas 1989), which is built on prosodic words. This is the structural correlate of phonetic invisibility. Members of the ‘PARSE’ family of constraints demand that the links in the prosodic hierarchy be established; let us use ‘M-PARSE’ for the constraint which requires the structural realization of morphological properties.

Applying this reasoning to the Latin case changes the game in certain crucial respects. Above we assumed that a lexical form /re/ inevitably had membership in a lexical category; consequently failure of phonological parsing led inevitably to violation of  $LX \approx PR$ . Now we assume that a form “has” a lexical category in the relevant sense only when the morphological parse is accomplished. A morphologically unassembled item *cannot* violate  $LX \approx PR$ , since the phonological string is not in the “is a” relation with a morphological category.

Under this assumption, it is not possible to rank FTBIN and  $LX \approx PR$  with respect to each other except arbitrarily, since they are both satisfied in optimal output. With M-PARSE ranked below them, the optimal output for subminimal input will lack morphological structure. To see how this plays out, examine the following tableau.

(63) Effect of Morphological Null Parse on {re, N}

Candidates	FTBIN	$LX \approx PR$	M-PARSE	(Ph)-PARSE
a. [ (ré) <sub>F,PrWd</sub> ] <sub>N</sub>	* !			
b. [ re ] <sub>N</sub>	<i>vac</i>	* !		*
c. (ré) <sub>F,PrWd</sub> , N	* !	<i>vac</i>	*	
d.  re, N	<i>vac</i>	<i>vac</i>	*	*

Hierarchical morphological structure, when present, is indicated by labeled brackets; phonological structure by labeled parentheses. The *ad hoc* indication *vac* marks vacuous satisfaction. Constraints which are not crucially ranked with respect to each other are separated in the tableau by dotted rather than solid lines. We assume that faithfulness constraints like FILL, which would be violated when the input is augmented to minimality, are ranked above M- and Ph-Parse, and we do not display augmented candidates; we consider the reverse ranking below.

Under these assumptions, any attempt to give /re, N/ a prosodic analysis violates FTBIN, as in (a) and (c). Morphological analysis without phonological analysis violates  $LX \approx PR$ , as in (b). This leaves only the Null Parse, phonological and morphological, which satisfies, vacuously, both FTBIN

and  $LX \approx PR$ . Since the Null Parse  $\langle re, N \rangle$  violates M-PARSE, we want M-PARSE subordinated in the hierarchy so that the issue can be decided by FTBIN and  $LX \approx PR$ .

On this view, then, the underlying form of an item will consist of a very incompletely structured set of specifications which constrain but do not themselves fully determine even the morphological character of the output form. These specifications must be put in relation, parsed into structure, in order to be interpretable. In mild cases of PARSE violation, bits of underlying form will not correspond to anything in the output: underlying consonants will not be realized, underlying long vowels will appear as short due to unparsed moras, and so on. But in the face of a battery of high-ranking structural constraints which the input is not suited to meet when faithfully mapped to the output, an entire Null Parse can be optimal. In this case, there is no interpretable output from the input form, and we have what amounts to absolute ill-formedness.<sup>30</sup>

It is worth taking a look at another line of attack, developed further in §9, that bears on the analysis of certain kinds of ill-formedness. The Null Parse is a kind of neutralization with  $\emptyset$ . We can also have neutralization with something tangible. Suppose that we have  $/A/, /B/ \rightarrow \Omega$  for two distinct inputs A and B. Suppose further that the input-output pairing  $(A, \Omega)$  incurs more serious violations than  $(B, \Omega)$  does. Then, under the natural assumption that the lexicon is chosen to minimize violations in the input-output mapping, and assuming that there is no other (e.g. morphological) reason to choose violation-prone A, one would be compelled to say that what underlies visible  $\Omega$  is actually B and not A. For example, in the Latin case, suppose (contrary to the assumptions just explored) that putative /re/ must always be parsed as  $[re:]$ . Then, if there are no other relevant repercussions of underlying vowel quantity, the surface form  $[re:]$  will be identified as /re:/ underlyingly.

To implement this approach, assume moraic structure and a correlated structural constraint FILL- $\mu$ . With FILL- $\mu$  appropriately subordinated in the hierarchy, unfilled moras can be posited in optimal forms under the compulsion of higher-ranked constraints. Assume further that unfilled moras are interpreted in the output as continuations of a tautosyllabic vowel, phonetically interpreted as length. For input /re/, the analysis  $r[e]_{\mu}[\ ]_{\mu} \equiv [re:]$  now becomes superior to the Null Parse. The output  $[re:]$  satisfies FTBIN and  $LX \approx PR$ , just like the Null Parse, but has the additional virtue of satisfying both the phonological and the morphological versions of the PARSE constraint. This state of affairs is portrayed in the following tableau:

(64) **Besting the Null Parse** when FILL is low

Candidates	FTBIN, $LX \approx PR$	M-PARSE	(Ph-)PARSE	FILL- $\mu$
$\left[ (r[e]_{\mu}[\ ]_{\mu})_F \right]_{PrWd}, N$				*
re, N	<i>vac</i>	* !	* !	

<sup>30</sup> The notion of absolute ill-formedness is taken up again in §9 below.

We now ask why, given [re:] in the observed world, the abstract learner would bother to posit underlying /re/ in the first place. If there is no adequate response, we can never have /re/ in the lexicon. Monomoraic forms are absent from the lexicon on this analysis not because they lead to uninterpretable output, but because the output they lead to is better analyzed as coming from something else. We might call this effect “occultation,” since the possible input /re/ is hidden, as it were, behind /re:/, and therefore inaccessible.

Mester (1992:19-23) provides evidence that this analysis is appropriate for actual Latin. The form of argument is due originally, we believe, to D. Stampe (1969, 1973, 1973/79) and has been reasserted independently by Dell (1973), Hudson (1974), Myers (1991) and probably others we are unaware of. We deal with related matters in the discussion of inventory theory below in §9.

Stampean occultation cannot always be appropriate. One common worry should perhaps be set aside: how do we determine what the occulting body actually *is* in cases with inadequate surface clues? For example, in the case at hand, in the absence of further, typically morphophonemic evidence, all that’s required is that /re/ come out as bimoraic; many epenthetic pathways are open. This issue can in principle be resolved by markedness theory or perhaps even left unresolved: anything, after all, will do. But there are still many circumstances, particularly in morphology, where it appears that there is simply no well-formed output from many kinds of input. (See McCarthy & Prince 1993:§7 for discussion.) The output from {violet, er} cannot be ‘bluer’ nor can the output from {write, ation} be ‘inscription’: these combinations yield nothing *tout court*. For such cases, the Null Parse is superior to all alternatives.

### 4.3.2 Nonfinality and the Laws of Foot Form: Extended Minimality Effects


Like so many languages, Latin bans monomoraic items because it respects both  $LX \approx PR$ , which *demand*s feet, and FTBIN, which bans *monomoraic* feet. Like many other languages, Latin also keeps stress off word-final syllables, by imposing NONFINALITY on its prosodic structure. Parsing of bimoraic monosyllables, in outright violation of NONFINALITY, is forced by the dominance of  $LX \approx PR$ , as we have seen. In a less obvious class of cases, NONFINALITY also conflicts with FTBIN.

Consider bisyllabic words shaped LL and LH like *áqua* and *ámo*. Under the standard view of extrametricality, these must be analyzed as  $L\langle L \rangle$  and  $L\langle H \rangle$ , that is, as the exact equivalent of #L#. The naive expectation, then, is that such words should be impossible rather than plentiful. Hayes (1991/1995) terms this state of affairs ‘the unstressable word syndrome’.

The solution to the #LL# case is already at hand, given what has been established above. We know from monosyllable behavior that  $LX \approx PR \gg$  NONFINALITY.  $LX \approx PR$  never forces binarity violations, as it would if it were ranked above FTBIN. In any total ranking of the constraints in the grammar, FTBIN must dominate  $LX \approx PR$ , and therefore, by transitivity, it must dominate

NONFINALITY as well.<sup>31</sup> From this it follows that /LL/ will be parsed [(LL)<sub>F</sub>]<sub>PrWd</sub>. The following tableau should make this clear:

(65) **Best Parse for /LL/** FTBIN, LX≈PR >> NONFINALITY

Candidates		FTBIN	LX≈PR	NONFINALITY
a.	.a.qua.		* !	
b.	(á.) <sub>F</sub> qua	* !		
c.	 (á.qua) <sub>F</sub>			*

In short, FTBIN and LX≈PR jointly force *all* bimoraic words to have a complete metrical analysis,<sup>32</sup> eliminating extrametricality from both #H# and #LL#.

The behavior of words shaped LH is more intriguing. If NONFINALITY could be completely ignored, as assumed above, we'd expect L(́), and thus *e.g.* \*amóː with final stress. The candidate parse (á)moː fails FTBIN, while incorrect \*a(móː) satisfies both FTBIN and LX≈PR, running afoul only of low-ranked NONFINALITY.

Forms like ámoː have two virtues which will allow us to rescue them from the oblivion in which forms like \*á repose. First, such words are *bisyllabic*: unlike true subminimals, they have enough substance to support a binary foot. The foot may not be beautifully formed, since (́H) makes a poor trochee but satisfies FTBIN nonetheless. Second, the exhaustively monopodic (ámoː) nonetheless displays a kind of nonfinality: the stress peak is indeed off the final syllable. With the notion of relative Harmony securely in hand, we can take advantage of these partial successes. The trochaic parse (́H) is equipped to succeed modestly on the constraints FTBIN and NONFINALITY, besting other analyses which fail them entirely. To implement this program of explanation, we need to refine our analyses of nonfinality and foot form.

<sup>31</sup> In terms of the empirically-necessary domination order, we have LX≈PR >> NONFINALITY from (c) vs. (a), and FTBIN >> NONFINALITY, from (c) vs. (b). But either ranking between LX≈PR and FTBIN will produce the same outcome, since there is a solution, namely (c), that satisfies *both* of them at the expense of NONFINALITY. With a grammar defined as a total ranking of the constraint set, the underlying hypothesis is that there is *some* total ranking which works; there could be (and typically will be) several, because a total ranking will often impose noncrucial domination relations (noncrucial in the sense that either order will work). It is entirely conceivable that the grammar should recognize nonranking of pairs of constraints, but this opens up the possibility of *crucial* nonranking (neither can dominate the other; both rankings are allowed), for which we have not yet found evidence. Given present understanding, we accept the hypothesis that there is a total order of domination on the constraint set; that is, that all nonrankings are noncrucial.

<sup>32</sup> This contradicts the theory of Steriade 1988, in which the word-final accent that appears in pre-enclitic position is attributed to extrametricality of the final syllable. But Mester 1992 offers good reasons to doubt the Steriade proposal.

The desired version of nonfinality simply puts together its forms from above in (45) and (53).

(66) NONFINALITY

No head of PrWd is final in PrWd.

The head of the PrWd is, immediately, the strongest foot  $\mathcal{F}$  dominated by the node PrWd. By transitivity of headship (the head of  $X'$  counting also as the head of  $X''$ ), the PrWd will also be headed by the strongest syllable in  $\mathcal{F}$ . In Latin, then, the PrWd has two heads, one inside the other. NONFINALITY is violated when *either* abuts the trailing edge of the PrWd; we assume that each violation counts separately. The candidate *\*a(móː)* manages to violate NONFINALITY at both levels: the head foot is final, as is the head syllable. The form (*ámoː*) is more successful: only the head foot is in final position. This is the result we want: the form  $\#(\acute{L}H)\#$  achieves a modest degree of success, since it keeps the main stress — the head of the head foot — out of final position.<sup>33</sup>

Foot form is already known to be determined by a composite of various principles. There must be a constraint which sets the rhythmic type at either iambic or trochaic; call this RHTYPE=I/T. Other factors regulate the well-formedness of various syllable groups, qua feet and qua bearers of iambic and trochaic rhythm (Hayes 1985, 1991/1995; McCarthy & Prince 1986; Prince 1990; Kager 1992abc; Mester 1992). For present purposes, we focus only on candidate principles that are relevant to the badness of (LH) as a trochee. One such is the Weight-to-Stress Principle of Prince 1990 (cf. ‘w-nodes do not branch’ of Hayes 1980).

(67) **Weight-to-Stress Principle (WSP).**

Heavy syllables are prominent in foot structure and on the grid.

By the WSP, the trochaic group ( $\acute{L}H$ ) is subpar because it puts a heavy syllable in a weak position.<sup>34</sup>

---

<sup>33</sup> The contrast with Kelkar’s Hindi, which has  $\#\acute{L}\acute{H}\#$  and  $\#\acute{L}\acute{L}\#$ , admits of several interpretations. The relation PK-PROM  $\gg$  NONFINALITY yields the first of these without elaboration, but when PK-PROM is not at issue, we’d expect the Latin pattern of antepenultimacy. The direct approach would redivide NONFINALITY into NONFIN( $\sigma'$ ) (45) and NONFIN(F')(53), each specifying whether the head syllable or the head foot was being regulated. The argument also depends on the kind of feet which are present in Hindi; if they are fundamentally iambic, then nothing need be said. If they are moraic trochees, we must split NONFINALITY and subordinate to EDGEMOST the version that refers to the head foot. Otherwise, we incorrectly predict antepenultimate rather than penultimate stress in light-syllabled words. The foot structure of languages like Hindi is, to say the least, incompletely understood.

<sup>34</sup> A different line of attack is also available for exploration, since the WSP is effectively the converse of a principle that we have seen active above in Berber and Hindi: PK-PROM is repeated here for convenience:


(i) Peak-Prominence (PK-PROM). Peak(x) > Peak(y) if  $|x| > |y|$ .

PK-PROM favors analyses in which positions of prominence are occupied by the heaviest syllables. By PK-PROM, the ( $\acute{L}H$ ) trochee is relatively bad because the counter-analysis  $L(\acute{H})$  has a better (heavier) peak. PK-PROM, in contrast to the WSP, says nothing directly about the occurrence of H in nonhead position. Complications arise in analyses where  $L(\acute{H})$  is ruled out by higher-ranking constraints, for example NONFINALITY, and we will not pursue this alternative here.

Since the trochee (LH) is inferior with respect to the WSP, it will be avoided in parses of syllable strings in favor of other structures that meet the constraint — *ceteris paribus*.

In Latin, NONFINALITY (66) renders the *ceteris imparibus*, as it were. We must have NONFINALITY >> WSP to force all final syllables to be foot-loose and non-prominent, regardless of their weight. With all the parts of the argument put together, the best parse [ámo:] emerges as in the following:


(68) Treatment of #LH# (Classical Latin) FTBIN, LX≈PR >> NONFINALITY >> WSP

Candidates	FTBIN	LX≈PR	NONFINALITY	WSP
a.  [ (ámo:) ]			*	*
b. [ a(mó:) ]			** !	
c. [ (á) mo:]	* !			*
d. amo:]		* !		*

The feet considered here are trochaic; it is fatal to advance an iambic parse in violation of the top-ranking but unmentioned constraint RHTYPE=T, since a better alternative, trochaic not iambic, always exists.

The ranking of NONFINALITY above WSP is part of the grammar of Latin. When the ranking is permuted, we get a language that looks more like Kelkar’s Hindi, for example English. It is well-known, due to the work of R.T. Oehrle (1972), that English bisyllabic nouns of the form LH are mostly end-stressed (*police*) while all others are mostly initially stressed (HH: *árgyle*; HL: *bándit*; LL: *éd-da*). (For some discussion, see Halle, 1973; Zonneveld, 1976; Liberman & Prince 1977; Hayes 1982, 1991/1995.) When WSP >> NONFINALITY, we get the effect of ‘extrametricality revocation’. The analysis runs exactly parallel to that of *ámo:* just presented in Tableau (68). Observe that *amó:* is the local winner under WSP and the local loser under NONFINALITY; were it the case that WSP dominated NONFINALITY, *\*amó:* would defeat *ámo:*. The *amó:*-like form *police* [p<sup>h</sup>əli:s] therefore optimal in the English-like system, as shown in the following tableau:

(69) Treatment of LH in an English-like system: ... WSP >> NONFINALITY

Candidates	FTBIN	LX≈PR	WSP	NONFINALITY
a. [ (pó)lice ]			* !	*
b.  [ po(líce) ]				**
c. [ (pó) lice ]	* !		*	
d. police		* !	*	

The Latin-style ranking of NONFINALITY above constraints relevant to foot form can have even more radical consequences. In many otherwise iambic languages, NONFINALITY forces bisyllables to be *trochaic*. Familiar examples include Choctaw, Southern Paiute, Ulwa, Axininca Campa<sup>35</sup>. In such cases, we have NONFINALITY  $\gg$  RHTYPE. In Southern Paiute we find, for example, qaʔíni ‘my necklace’ but kúna ‘fire’. (Final syllables are voiceless; examples from Hale (undated).) Southern Paiute phonology is such as to make it clear that the effect occurs in all forms: NONFINALITY is therefore not restricted to a single head-foot of the prosodic word but applies to all feet and to their heads.

(70) **Rhythmic Reversal** due to NONFINALITY

Southern Paiute /puNpuNkuɲwɪtaɲwa/ ‘our (incl.) horses owned severally’

Candidates	NONFINALITY	RHTYPE=I
☞ (pumpúɲ)(kuɲwɪ)(taɲwa)		*
(pumpúɲ)(kuɲwɪ)(taɲwà)	* !	

The consequences are most dramatic in bisyllables, where the placement of the only stress in the word is affected:

(71) **Rhythmic Reversal** due to NONFINALITY: Southern Paiute /kuna/ ‘fire’

Candidates	NONFINALITY	RHTYPE=I
☞ (kúna)		*
(kuná)	* !	

We have only sketched the core of the phenomenon here. For further discussion, see McCarthy & Prince 1993:§7;<sup>36</sup> detailed further exploration is found in Hung (1993).

<sup>35</sup> Relevant references include: Choctaw (Muskogean)— Lombardi & McCarthy 1991; Munsee (Algonkian)— Hayes 1985; Southern Paiute (Uto-Aztecan)— Sapir 1930; Ulwa (Misumalpan)— Hale & Lacayo-Blanco 1988; Axininca (Arawakan) Payne 1981.

<sup>36</sup> McCarthy & Prince note that in an iambic system operating under the compulsion of exhaustive parsing (PARSE- $\sigma$ , i.e. syllables into feet), the appearance of Left-to-Right parsing is a consequence of NONFINALITY. In odd parity strings, NONFINALITY puts the free syllable at the very end, no matter where it is in the ranking, above or below RHTYPE. (In even parity strings, the domination relationship determines whether the iambic foot will be inverted at word-end.) This result holds whether the iamb is sensitive to quantity or not. It may be relevant that almost all iambic systems are Left-to-Right in directional sense (see Kager 1992c for discussion from a different perspective).



#### 4.4 Summary of Discussion of the *Except When* Effect

Explicating the ‘except when’ or blocking pattern in terms of constraint domination has led to new understanding of a variety of prosodic phenomena. We have been able to put edge-oriented infixation on a principled basis, leading to predictions about the kind of affixes liable to such treatment. The properties previously attributed to a formal theory of extrametricality follow from construing extrametricality as an interaction effect which arises when items placed by the gradient constraint EDGEMOST are subject to higher-ranking substantive constraints such as NONFINALITY or ONS which force violations of EDGEMOST, minimal of course. Effects of ‘extrametricality revocation’, rather than being due to quirks in the formal theory or special re-write rules applying in arbitrary environments, are seen instead to follow from the dominance of constraints such as FTBIN, PK-PROM, and LX≈PR, easily recognizable as authentic, independently-required principles of prosody and of the prosody/morphology interface.

The interaction that ‘revokes extrametricality’ is exactly of the same sort as that which introduces it in the first place — constraint domination. Just as NONFINALITY overrules EDGEMOST producing the appearance of extrametricality, so does FTBIN (for example) overrule NONFINALITY, taking extrametricality away. We have examined a number of such cases, embodying the *except when* or *blocking* form of descriptive generalization.

- Syllables are descriptively extrametrical everywhere in a language like Latin *except when* nonparsing would leave nothing (monosyllables) or a single mora (bisyllables /Lσ/).

- Feet take on their independently favored forms — obeying the WSP and the language’s RHTYPE — everywhere *except when* in bisyllables, where FTBIN is at stake due to the force of NONFINALITY.

- Stress falls on an edgemost syllable *except when* there is a heavy syllable in the word (in which case it falls on the edgemost such syllable).

- Stress falls on a nonfinal syllable *except when* it is the heaviest syllable in the word.

- Affixes are situated at the edge of the {affix, stem} collocation *except when* stem-internal positioning results in more harmonic syllable structure (and even then, given satisfaction of the syllabic requirements, they are positioned at minimal distance from the target edge.)

The sometimes intricate patterns of dependency that lie behind descriptive notions like *minimality* and *extrametricality* thus emerge from the interaction of principles of prosodic form. The success of the theory in rationalizing this domain and opening up the way to new forms of explanation stands as a significant argument in its favor.

#### 4.5 Except meets Only: Triggering and Blocking in a Single Grammar<sup>37</sup>

To conclude our overview of generalization patterns, we take on some empirical issues made accessible by the results of the preceding discussion. We show how nonfaithful parsing interacts with the kind of minimality, extrametricality, and prominence effects just examined. The empirical focus will be on Latin, viewed with finer resolution of detail, and we will show how a renovated

---

<sup>37</sup> Thanks to Armin Mester for discussing various points with us. Errors are ours, of course.

conception of the language's prosody is mandated. By its nature, the analysis must reach a fair level of complexity, and the reader who wishes to focus on the main line of the argument may wish to return after examining §6.

Words ending in difficult-to-foot sequences like LH pose a challenge that is sometimes met quite aggressively. In the pre-classical Latin of Terence and Plautus — presumably reflecting a less normative interpretation of the actual spoken language — bisyllabic words /LH/ may come out optionally as LL, whence *ámo* instead of *ámo*Ꞥ, yielding a far more satisfactory trochee, one that does not violate WSP. This phenomenon goes by the name of 'iambic shortening', in reference to the quantitative shape of the input, and has recently been examined in number of studies, especially Mester 1992, whose results have inspired these remarks, but also including Allen 1973; Devine & Stephens 1980; Kager 1989; Prince 1990; Hayes 1991/1995. Mester 1992 emphasizes, with Devine & Stephens, that words shaped /...HLH/ suffer an entirely parallel reduction to ...HLL — also optionally, and at the same historical period. Here are some examples involving the 1<sup>st</sup> person singular present ending -ō:<sup>38</sup>

(72)	/HLH/	/LLH/	/HH/
	dēsinō	studeō	laudō
	nesciō	habitō	mandō

This is called 'cretic shortening' after the cretic foot of Greek lyric meters (HLH). Mester produces a wide variety of phenomena of Latin phonology and morphology, including these shortening effects, which are responsive to constraints demanding exhaustive footing with high-quality trochees. The structure (H)(LL) puts all syllables into bimoraic trochees, where (H)L(H) traps a syllable between two feet, (H)LH leaves the final two syllables dangling, and (H)(LH) involves a poor final trochee, the very same (LH) that is forced on bisyllables like *ámō* in the classical language. Common to both iambic and cretic shortening is the interpretation of underlying ...LH# as surface ...LL#, which results in notably improved prosodic organization.

Taking this explanation seriously will push our understanding of Latin prosodic organization well beyond what can be deduced from the simple facts of main-stress placement. For one thing, words HLL must be structured (H)(LL) not (H)LL, as has been assumed in previous scholarship, and words HH must be structured (H)(H), not (H)H. Mester sees a two-step process: standard foot-assignment and main-stressing, followed by a stage of repair in the direction of well-formedness, which incorporates loose syllables into foot structure when possible, with the assistance of a rule 'Remove μ'. By contrast, and in line with the general thrust of our argument throughout, we wish to explore the idea that iambo-cretic shortening is a direct by-product of the one basic parse, in the same way that epenthesis and deletion are by-products of the one syllabic parse (Selkirk 1981, Itô

---

<sup>38</sup> Glosses are dēsinō 'cease', studeō 'strive', laudō 'praise', nesciō 'don't-know', habitō 'reside', mandō 'entrust', all 1psg. We have normalized Mester's example mandā 'entrust 2sg. imp.' to the 1psg. Note that in later Latin there is a phenomenon whereby word-final o is often short, regardless of preceding context (Mester 1992:31 fn. 43). This phenomenon is irrelevant to the Pre-Classical situation, but could mislead the casual observer.

1986, 1989). On this view, iambocretically shortened forms are among the many candidate analyses that forms  $\sim$ HLH# and #LH# are subject to in any language. The constraint system of Pre-Classical Latin, at least at its colloquial level, picks them out as optimal.

Before grappling with the details of iambocretic shortness, it is useful to map the structure of the basic rhythmic system of the language. This involves three components:

- a. Positional theory: nonfinality and edgemostness.
- b. Parsing theory: to compel footing of syllables, and syllabification of segments and moras.
- c. Foot theory: of trochees, their shape and proclivities.

First, the positional constraints. Iambo-cretic shortening indicates that Latin final syllables are not to be excluded in principle from foot structure; quite the contrary. The doctrine of total extraprosodicity is a superficial conclusion drawn from focusing on examples like (*spátu*)*la*, where there is simply nothing to be done. We know that /dēsīn-ō/ will be parsed (dē)(sino) in the colloquial language, and from this we can infer that *blandula* is parsed (blan)(dula) and that *mandō* is (man)(dō). Maximality of prosodic organization is not abandoned wholesale just to place the main-stress in the right position. (It is abandoned minimally, as we expect.) There is no law banning feet from final position; rather it is the prosodic heads of the word — main stress and main foot — that are banned. This is exactly what was claimed above in the statement of NONFINALITY for Latin (66), which we repeat here:

(73) **NONFINALITY**

No prosodic head of PrWd is final in PrWd.

The constraint applies to the main-stressed syllable itself and to the foot in which it is housed. Descriptively, one would say that main-stress falls not on the *last* foot, as in previous analyses, but on the last *nonfinal* foot, rather like Palestinian Arabic, Cairene Radio Arabic, and Munsee, among others (Hayes 1991/1995).

The constraint of NONFINALITY pushes the main stress off the final syllable, often leaving open a number of possibilities for its location. As usual, the major positioning constraint is EDGEMOST: the prosodic heads fall as far to the right as possible. We must have, of course, NONFINALITY  $\gg$  EDGEMOST, otherwise no nonfinality effects would be observable.<sup>39</sup> The constraint EDGEMOST( $\varphi$ ;E) penalizes forms according to the extent that item  $\varphi$  is removed from edge E. In Latin  $\varphi$  = Prosodic head, E = right edge. For purposes of reckoning distance from the edge, we will take the relevant prosodic head to be the main-stressed syllable itself, and the distance from the edge will be measured in syllables.

To see the effects of this scheme, consider first forms LLL, which have two reasonable parses ( $\acute{L}L$ )L and L( $\acute{L}L$ ). For head foot, we write F', for head syllable of head foot we write  $\sigma'$ .

---

<sup>39</sup> This is an instance of Pāṇini's Theorem, §5 below, whereby with relations between constraints A and B, the ranking B  $\gg$  A entails that A has no effect on the grammar. Notice that we are not saying that they are 'ranked by the Elsewhere Condition': they are ranked by the facts; either ranking is possible.

(74) **Nonfinality Effect** in trisyllables

Candidates	NONFINALITY(F',σ')	EDGEMOST(σ',R)
☞ (spátu)la		tula#
spa(túla)	*F'	la#

Assumed is the dominance of FTBIN and of LX≈PR, which entails that a form must have a least one foot and a binary one at that. In the more general context, the effects of a principle of exhaustive metrical analysis, familiar from the earliest work in the area (Lieberman 1975), is visibly at work. This principle is part of the parsing theory, and we will call it PARSE-σ, omitting from the name the information that σ is parsed into F. It ensures, for example, that forms shaped HH will be parsed (H́)(H) rather than (H́)(H). Both candidates have perfect nonfinality of F' and σ' and both agree in edgemostness; only PARSE-σ dictates that the second foot must be posited. Similarly, forms LLH must be parsed (ĹL)(H). That PARSE-σ is relegated to a subordinate role becomes apparent when we examine forms LLLL. Examination of the natural candidates for analysis of LLLL shows that EDGEMOST must dominate PARSE-σ:

## (75) EDGEMOST &gt;&gt; PARSE-σ, from LLLL

Candidates	NONFINALITY(F',σ')	EDGEMOST(σ',R)	PARSE-σ
a. ☞ pa(tríci)a		i.a#	**
b. (pátri)(ci.a)		tri.ci.a# !	
c. (patri)(cí.a)	*F'		

The correct output *patricia* (a) is poorly σ-parsed, while both of its competitors are perfect on that score. EDGEMOST, however, rules in its favor in the competition with (b), and so must dominate PARSE-σ. Note that the pre-antepenultimate form (b) is correct for Palestinian Arabic, suggesting a different ranking of the constraints in that language.

These considerations give us the core of the system of both Classical and Pre-Classical Latin. We have *Position >> Parsing*, or, in detail:

(76) **Antepenultimacy**

NONFINALITY(F',σ') >> EDGEMOST(σ',R) >> PARSE-σ

The fundamental constraints LX≈PR and FTBIN are of course superordinate.

On this account, a word LLL is analyzed (ĹL)L not because the final syllable is marked extrametrical, but rather because the main *foot* stands in a nonfinal position, a pure nonfinality effect. The fate of words LLLL is decided by a combination of NONFINALITY, which winnows the candidate set down to L(ĹH)L and (ĹL)(LL), and EDGEMOST, which decides the matter in favor of forms in which the prosodic head stands nearest to the end of the word.

This kind of argument parallels the one given above about prefixing infixation (§4.1), and can only be made in a theory that compares across the set of possible structures. Under Bottom-Up Constructionism, which builds structure by deterministic rule, the feet must be fixed in place first, and you can only choose the head-foot from among the ones you have before you.

Exactly a *single* syllable ends up extrametrical here. But this follows from the very interaction of EDGEMOST and NONFINALITY which is necessary to generate the extrametricality effect in the first place. There is no special notion [+extrametrical], distinct from ‘unparsed’ or ‘occupying weak position in prosodic structure’, conditions which befall many an element, for many reasons. Consequently, as argued above in §4.3, there is no theory of extrametricality as a formal device, which requires among its axioms a principle limiting the scope of the feature [+extrametrical] to a single constituent. Because violations are always minimal, only the minimal amount of non-edgemostrness will be tolerated. This minimum will often turn out to be a single constituent. (But not necessarily so, unless other factors conspire: cf. Tagalog *gr-um-adwet*, §4.1, where *gr* need not be regarded as constituent for the explanation to go through.)

The remaining component of the system is the foot theory. The bedrock constraints are FTBIN and RHTYPE=T. These are unviolated in the optimal forms of the language: every foot is binary on syllables or moras, and every foot is trochaic. In addition, universal prosody recognizes the Weight-to-Stress Principle WSP, which urges that the intrinsic prominence of heavy syllables be registered as prosodic prominence; this discriminates against the quantity-prominence mismatch of trochees (LH). Feet (HL) satisfy all these constraints but are known to be marked or even absent in trochaic systems (Hayes 1987, Prince 1990; Mester 1992); we wish to ban these on grounds of *rhythmic* structure, which favors length at the end of constituents. Let us call the relevant constraint ‘Rhythmic Harmony’ (RHHRM); for present purposes we can simplify its formulation to \*(HL).

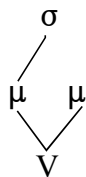
Of these constraints, only the WSP shows any mobility in the dialects (or register levels) of Latin under discussion. In Classical Latin, as shown above in (68), we must have NONFINALITY >> WSP, to obtain (ámo:) rather than \*a(mó:). Both candidates contain a foot, satisfying the morphology/prosody interface constraint LX≈PR; both contain binary feet, satisfying FTBIN; but only (ámo:) succeeds in getting the main stress off the last syllable. The cost is violation of the WSP. In the iambocretic-shortening register of Pre-Classical Latin, by contrast, we have (ámo) and no violation of the WSP at all. All four foot-relevant constraints are unviolated, and are therefore not dominated crucially by any other constraints in the grammar. These observations are summarized in the following table.

## (77) Foot Form Constraints

Constraint	Effect	Status
FTBIN	$F = \mu\mu, \sigma\sigma$	unviolated
RHTYPE=T	$(\sigma\sigma) = (\acute{\sigma}\sigma)$	unviolated
RHHRM	* (HL)	unviolated
WSP	* Ĥ	Violated in Classical Latin. Unviolated in Pre-Classical shortening register.

Having secured the proper infrastructure, we can approach the issue of iambocretic shortening. Mester proposes a special rule of repair, ‘Remove  $\mu$ ’, which works in concert with his second round of footing. Optimality Theory does not recognize repair strategies as distinct from the ordinary resources available for assigning structure in the first place. Structure is posited freely, though always subject to evaluation. Empty nodes are violations (of the constraint we have called FILL), but violation can be forced. Similarly, elements present in the input can remain structurally unanalyzed, violating PARSE. Such violations are tolerated in an optimal candidate only when they lead to better performance on higher-ranked constraints. In the case at hand, an underlying mora (as in the second syllable of /amo:/) is left unparsed — unattached to syllable structure — and hence uninterpreted, giving rise to a phonetic short vowel. The resulting structure looks something like this:

## (78) Syllabically Unparsed Mora



This is a monomoraic syllable. Failure to construe the second mora is a violation of PARSE, specifically of the constraint PARSE- $\mu$ , which must be distinguished from the other members of the PARSE-*element* family. In Classical Latin, quantity is stable under prosodic analysis; PARSE- $\mu$  is widely observed and no constraint discussed here forces it to be violated. Therefore, PARSE- $\mu$  in Classical Latin is undominated by the constraints at hand.<sup>40</sup> But in the shortening register of Pre-Classical Latin, PARSE- $\mu$  is subordinated in the ranking, leading to the appearance of phonetic short vowels as the globally optimal but locally unfaithful renditions of lexical long vowels.

Following Mester 1992: 33, we want /HLH/ words like /di:cito:/ ‘say, imp. fut.’ to be parsed (di:)(cito)⟨:⟩. Similarly, we want /LH/ words like /amo:/ to be parsed (amo)⟨:⟩. With the ad hoc

<sup>40</sup> By contrast, numerous violations of PARSE-*segment* are forced by syllable structure conditions, as in /mell/ → mel, /cord/ → cor, /ment+s/ → mens (Steriade 1982). See below §6-8 for further discussion.

notation  $H^-$  used to indicate a potentially heavy syllable parsed as light, as for example in diagram (78), the relevant lexical-surface associations can be written as follows:


(79) **PC Latin Parsing of Interesting Strings**

Base	Faithful	*PARSE- $\mu$
a. HH	(H)(H)	
b. LH		(LH $^-$ )
c. HLL	(H)(LL)	
d. HLH		(H)(LH $^-$ )

What forces the PARSE- $\mu$  violations? For bisyllabic words, it is clear that the parse (LH $^-$ ) successfully avoids the wretched trochee (LH), which violates the WSP. Therefore we have WSP  $\gg$  PARSE- $\mu$ , meaning that it is more important to ensure that all heavy syllables are stressed than to retain underlying vowel length. One way to satisfy the WSP is to omit potential heaviness from syllables that stand in unstressable positions.

A second constraint that crucially dominates PARSE- $\mu$  is PARSE- $\sigma$ , which compels exhaustive footing. This relation is manifest in the treatment of forms ending in HLH, which show cretic shortening.

(80) **PARSE- $\sigma$   $\gg$  PARSE- $\mu$**  from /HLH/

Candidates	PARSE- $\sigma$	PARSE- $\mu$
 (H́)(LH $^-$ )		*
(H́)L(H)	* !	

The parse \*(H́)L(H) has nothing wrong with it but the ‘trapping’, as Mester puts it, of the middle syllable: it is unaffiliated with F. The counter analysis (H́)(LH $^-$ ) is fully  $\sigma$ -parsed, but the price is leaving the final  $\mu$  out of syllable structure. This shows that PARSE- $\sigma$  is dominant, forcing the PARSE- $\mu$  violation. This ranking recognizes Mester’s basic claim: that exhaustive parsing is the motive force behind the shortness effect.

With PARSE- $\mu$  subordinated in the ranking, feet (H $^-$  L) also become significant contenders. They also resolve the parsing problem presented by HLH, leading to \*(*dési*)(*no:*), for example, in place of actual (*dé:*)(*sino*). Both candidates are fully  $\sigma$ -parsed; both have nonfinal prosodic heads; and they agree on the rightmostness of the mainstress. The only difference is the location of the H that is incompletely parsed as H $^-$ . The outcome (H $^-$  L)H is witnessed in various languages, famously English (Myers 1987; Prince 1990) and Boumaa Fijian (Hayes 1991/1995), but not Latin. What makes the difference in Latin? The plausible candidate is PK-PROM, which we repeat from (37):

(81) **Peak-Prominence (PK-PROM).**

Peak(x)  $\succ$  Peak(y) if  $|x| > |y|$

In terms of a two-way H/L contrast, this means that  $\acute{H}$  is a better peak than  $\acute{L}$ . Because the unparsed mora of  $H^-$  makes it into a light syllable, the foot ( $H^-L$ ) has inferior peak-prominence compared with ( $\acute{H}$ ). The contrast becomes important in the analysis of HLH, as this tableau demonstrates:

(82) PK-PROM and PARSE- $\mu$ 

Candidates	PK-PROM	PARSE- $\mu$
$\rightarrow$ ( $\acute{H}$ )(LH $^-$ )	H	*
( $\acute{H}^-$ L)(H)	H $^-$  = L  !	*

This is not a ranking argument: PARSE- $\mu$  does not discriminate between these candidates, so the constraints are not in diametric conflict. PK-PROM will decide the issue no matter where it is located in the hierarchy.<sup>41</sup> In Latin PK-PROM cannot be allowed to play the same role that it does in Hindi, singling out a heavy syllable anywhere among a string of light syllables; the positional constraints must dominate it, EDGEMOST in particular:

(83) EDGEMOST  $\gg$  PK-PROM, from HLLL /incipere/ ‘to begin’

Candidates	EDGEMOST	PK-PROM
$\rightarrow$ (in)( <b>c</b> ipe)re	pe.re#	L
( <b>in</b> )(cipe)re	ci.pe.re# !	H

This relation also holds in the grammar of Classical Latin, where PK-PROM has no visible effects. It is worth emphasizing that PK-PROM is quite distinct notionally from the WSP. The WSP goes from weight to stress: ‘if heavy then stressed’ (equivalently, ‘if unstressed then light’). PK-PROM essentially goes the other way: ‘if stressed then heavy’ (contrapositively and equivalently: ‘if light then unstressed’). To see this, imagine PK-PROM playing out over a binary weight contrast; it says  $\acute{H} > \acute{L}$  or in the language of violations,  $*\acute{L}$ . The WSP has nothing to say about this configuration. These two halves of the purported biconditional ‘heavy iff stressed’ have very different status in stress systems, as Prince 1990 observes; the present case is a further demonstration of this fact.

We have now established the form of the constraint system. For convenience of reference, the rationale for the crucial rankings is summarized in the following table:

<sup>41</sup> Observe that PK-PROM as stated deals only with the main stress, the peak of the PrWd. Feet (HL) are banned everywhere in the word, and so the explanation must be extended to these cases. The issue does not arise to the right of the main stress, and we will return to it only after the main argument is laid out.



## (84) Support for Rankings

Ranking	Because:	Remarks
NONFINALITY(F',σ') >> EDGEMOST(σ';R)	( <u>́</u> LL)L ] > L( <u>́</u> LL) ]	<i>Shared, all Registers</i> Position of main stress
EDGEMOST(σ';R) >> PARSE-σ	L( <u>́</u> LL)L ] > ( <u>́</u> LL)(LL) ]	
EDGEMOST(σ';R) >> PK-PROM	H <u>́</u> LLL ] > <u>́</u> HLLL ]	
WSP >> PARSE-μ	[( <u>́</u> LH <sup>-</sup> ) ] > [( <u>́</u> LH) ] ( <u>́</u> )(LH <sup>-</sup> ) ] > ( <u>́</u> )(LH) ]	<i>Shortening Register:</i> Iambocretic effect
PARSE-σ >> PARSE-μ	( <u>́</u> )(LH <sup>-</sup> ) ] > ( <u>́</u> )L(H) ]	

Observe that the three relations above the double line are shared in all dialects or speech-levels; those below it delimit the characteristics of iambocretic shortening.

To conclude this discussion, we assemble the constraint relations established for Latin and go on to demonstrate the efficacy of the constraint system in dealing with the key examples.

## (85) Constraint Structure of PC Latin Shortening Grammar

## a. Undominated

WSP, RHHRM, FTBIN, RHTYPE=T, LX≈PR

## b. Main Sequence

LX≈PR, FTBIN >> NONFINALITY(F',σ') >> EDGEMOST(σ';R) >> PARSE-σ >> PARSE-μ

## c. Weight Effect

WSP >> PARSE-μ

## d. Bounding

EDGEMOST(σ';R) >> PK-PROM

Empirical considerations do not force a total order on the set of constraints. Any total order consistent with the partial order will yield equivalent results. Without violating the crucial rankings we can place all four of the foot-relevant constraints WSP, RHHRM, FTBIN, and RHTYPE=T in a single package at the top of the hierarchy, which we will label 'foot form'. More arbitrarily, under the compulsion of the planar format of the tableau, we will list PK-PROM at the very bottom. This divides the constraint system into four blocks:

## (86) Block Structure of the Constraint system, in domination order

## a. Foot Form

## b. Position


## c. Parsing

## d. Prominence

To simplify the presentation, all candidate parses that fail to satisfy LX≈PR and FTBIN will be omitted from consideration, since violation is obvious and inevitably fatal.


First, let us contrast the treatment of bisyllables LH and HH.

**(87) Parsing of LH**

Candidates	Foot Form	Position		Parsing		PK-PROM
		NONFINAL	EDGEM	PARSE- $\sigma$	PARSE- $\mu$	
/LH/						
a.  (LH <sup>-</sup> )		*F'	$\sigma\#$		*	L
b. (LH)	*WSP!	*F'	$\sigma\#$			L
c. L (H)		*F' * $\sigma'$ !	$\emptyset\#$	*		H


Here the parse (LH<sup>-</sup>) is optimal because (LH) is a poor foot violating WSP, and because (LH) violates both requirements on the nonfinality of heads. This shows that we have ‘iambic shortening’.

**(88) Parsing of HH**

Candidates	Foot Form	Position		Parsing		PK-PROM
		NONFINAL	EDGEM	PARSE- $\sigma$	PARSE- $\mu$	
/HH/						
a.  (H)(H)			$\sigma\#$			H
b. (H)H	*WSP!		$\sigma\#$	*		H
c. (H)H <sup>-</sup>			$\sigma\#$	*!	*	H


The form (H)(H) satisfies all constraints maximally well and its rivals must all do something worse.

## (89) Parsing of HLH

Candidates /HLH/	Foot Form	Position		Parsing		PK-PROM
		NONFINAL	EDGEM	PARSE- $\sigma$	PARSE- $\mu$	
a.  (H́)(LH <sup>-</sup> )			$\sigma\sigma\#$		*	
b. (H́L)(H)			$\sigma\sigma\#$		*	* !
c. (H́)L(H)			$\sigma\sigma\#$	* !		
d. (H́)(LH)	*WSP !		$\sigma\sigma\#$			
(H́L)(H)	*RHHRM !		$\sigma\sigma\#$			

Perhaps the most interesting comparison is between (H́)(LH<sup>-</sup>) and (H́)L(H). It's decided by PARSE- $\sigma$ , which forces the inclusion of syllables in foot structure, at the expense of an unfaithful rendering of the underlying moraic structure.

## (90) Parsing of HLL

Candidates /HLL/	Foot Form	Position		Parse		PK-PROM
		NONFINAL	EDGEM	PARSE- $\sigma$	PARSE- $\mu$	
a.  (H́)(LL)			$\sigma\sigma\#$			
b. (H́ L) L	*RHHRM !		$\sigma\sigma\#$	*		
c. (H)(ĹL)		*F' !				

Like its moraic parallel HH, the wordform HLL has the near-perfect parse (H́)(LL), with only EDGEMOST being violated, a necessary infraction. Its rivals must all incur more serious violations.

This completes our review of the basic patterns. A couple of further subtleties deserve mention. Iambocretic shortening not only affects long vowels, it also treats *closed syllables* as light, as evidenced by versification practice (Allen 1973, Kager 1989, Mester 1992). We have quantitative analyses like these:<sup>42</sup>

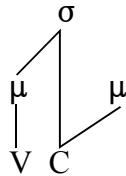
## (91) Light Closed Syllables

Base	Analysis	Pattern
a. /kanis/	(.ka.nis.) <sub><math>\mu\mu</math></sub>	(LH <sup>-</sup> )
b. /volupta:te:z/	(.vo.lup.) <sub><math>\mu\mu</math></sub> (ta:) <sub><math>\mu\mu</math></sub> (te:z) <sub><math>\mu\mu</math></sub>	(LH <sup>-</sup> ) (H) (H)
c. /ve:nerant/	(ve:) <sub><math>\mu\mu</math></sub> (.ne.rant.) <sub><math>\mu\mu</math></sub>	(H) (LH <sup>-</sup> )

<sup>42</sup> Glosses: 'dog'; 'desires'; 'come' 3pl.fut.pf. Examples from Mester 1992:12;31.

In such cases, the syllable-closing consonant is analyzed as nonmoraic, merely adjoined to the syllable, so that .nis. = .n[i]<sub>μ</sub>s, for example (cf. Kager 1989). What is ineffective, here, is the parsing principle requiring that a syllable-final consonant (sequence) correspond to a heaviness-inducing weak mora. To ensure the parallelism with long-vowel shortening, we can assume that moraic parsing is not itself violated, but rather that the mora it licenses remains unparsed syllabically. The structure looks like this:

(92) **Unparsed Syllable-Closing Mora**



If this kind of structure is admitted, both kinds of lightness effects of iambo-cretic shortening show up as PARSE- $\mu$  violations.<sup>43</sup>


Iambic shortening also shows up in non-final position, as example (91b) illustrates.<sup>44</sup> The following tableau demonstrates that the constraint hierarchy already predicts this outcome:

---

<sup>43</sup> These phenomena can also be construed in metrical rather than syllabic terms, perhaps more interestingly. Suppose that the actual terminal nodes of metrical structure are the positions on the first grid row (Prince 1983, Halle & Vergnaud 1987). (Bracketing of syllable strings is derivative from this in the obvious way.) Mapping syllable structure to the grid is what's at issue. In general,  $\mu \rightarrow x$ . But we can also have  $[\mu\mu]_{\sigma} \rightarrow x$ , "compression" in the manner of Yupik (Woodbury 1981, Hewitt 1992), somewhat like the "mora sluicing" of Prince 1983. The constraint that's being violated is not PARSE- $\mu$ -to- $\sigma$  but rather PARSE- $\mu$ -to- $x$ . To make it perfectly parallel to discussion in the text, instead of  $\mu_i\mu_j \rightarrow x_ix_j$ , we get  $\mu_i\mu_j \rightarrow x_i$ , with  $\mu_j$  unassociated. Thus both V: and VC violate PARSE- $\mu$ -to- $x$  equally. Compressed V: is interpreted as equivalent to a short vowel; that is, quantity is read off the first layer of the grid. Evidence for this view comes from the fact that geminates may suffer iambic shortening, e.g. *supellectilis* 'utensils' without apparently being literally degeminated (Mester 1992: 17n).

<sup>44</sup> Mester (1992: 17n) points to a number of other subtleties, such as the fact that the vowel-shortening wing of the phenomenon applies only sporadically to internal long vowels. He suggests that preserving lexical (distinctive) quantity may drive this difference in behavior. We refer the reader to Mester's discussion.

## (93) Internal Nonparsing of Moras

Candidates	Foot Form	Position		Parsing		PK-PROM
		NONFINAL	EDGEM	PARSE- $\sigma$	PARSE- $\mu$	
LHHH						
a.  (LH <sup>-</sup> )(H́)(H)			$\sigma\#$		*	H
b. L(H)(H́)(H)			$\sigma\#$	* !		H

An interesting issue arises in longer words which have the potential for feet (H<sup>-</sup>L) before the main stress, for example HLHH. As noted in fn. 41, p. 66 above, if PK-PROM is stated so as to apply only to the peak or main stress of the PrWd, it will have no consequences outside the head foot. Without further remark, we expect (H<sup>-</sup>L)(H́)(H) from HLHH. This avoids the Foot Form violation (HL), which we know to be avoided in the parse of /HLH/, shown in (89); and it achieves complete foot-parsing at the expense of PARSE- $\mu$ , which is well-motivated in the grammar. We suggest that PK-PROM in fact applies to all “peaks” or heads, both of PrWd and of F. But within this subsystem, there is also a ranking of priorities: evaluation of the head of PrWd takes priority over evaluation of foot-heads. To see this, consider the rival parses \*(H́<sup>-</sup>L)(H) and (H́)(LH<sup>-</sup>). At the foot level each sports one heavy head and one light head, since  $|H<sup>-</sup>| = |L|$ . It is only at the PrWd level that the decision can be made in favor of the correct form, which has a heavy main peak. In §8, we provide a general mechanism for constructing complex constraints like PK-PROM from the coordination of two distinct scales of relative prominence, here (1) head of PrWd vs. foot-head and (2) heavy syllable vs. light syllable.

A final, fundamental question raised by the analysis concerns the relation between Classical Latin and Pre-Classical Latin, or, more precisely, between the formal register and the informal shortening register of the Pre-Classical language. The shortening register is derived by two re-rankings: PARSE- $\mu$  is made subordinate and the WSP rises to the top. It might be objected, on aesthetic grounds, that one should aim to get the difference out of a single re-ranking. This seems to us too superficial a judgment. The notion of colloquial register surely has a substantive characterization which is not even approximated by counting changes in a constraint or rule system; certain single re-rankings will massively alter the surface appearance of the system. In Latin, the effect of the re-ranking is to render the foot-defining block of constraints entirely transparent, at the cost of some failures to realize underlying quantity. This is consistent with the general sense that colloquial language simplifies prosodic structures, rendering them in closer accord with universal *structural markedness* constraints, while subordinating *Faithfulness*.



## 5. The Construction of Grammar in Optimality Theory

Phonological theory contains two parts: a theory of substantive universals of phonological well-formedness and a theory of formal universals of constraint interaction. These two components are respectively the topics of §5.1 and §5.2. Since much of this work concerns the first topic, the discussion here will be limited to a few brief remarks. In §5.3, we give Pāṇini's Theorem, a theorem about the priority of the specific which follows from the basic operation of Optimality Theory as set out in §5.2.

### 5.1 Construction of Harmonic Orderings from Phonetic and Structural Scales

To define grammars from hierarchies of well-formedness constraints, we need two distinct constructions: one that takes given constraints and defines their interactions, another that pertains to the constraints themselves. The first will be discussed at some length in §5.2; we now take up the second briefly.

Construction of constraints amounts in many ways to a theory of contextual markedness (Chomsky & Halle 1968: ch. 9, Kean 1974, Cairns & Feinstein 1982, Cairns 1988, Archangeli & Pulleyblank 1992). Linguistic phonetics gives a set of scales on phonetic dimensions; these are not well-formedness ratings, but simply the analyses of phonetic space that are primitive from the viewpoint of linguistic theory. (We use the term 'scale' in the loosest possible sense, to encompass everything from unary features to n-ary orderings.)

Issues of relative well-formedness, or markedness, arise principally when elements from the different dimensions are combined into interpretable representations. High sonority, for example, does not by itself entail high (or low) Harmony; but when a segment occurs in a structural position such as nucleus, onset, or coda, its intrinsic sonority in combination with the character of its position gives rise to markedness-evaluating constraints such as HNUC above. Similarly, tongue-height in vowels is neither harmonic nor disharmonic in isolation, but when the dimension of ATR is brought in, clear patterns of relative well-formedness or Harmony emerge, as has been emphasized in the work of Archangeli & Pulleyblank (1992). These *Harmony scales* are intimately tied to the repertory of constraints that grammars draw on. Inasmuch as there are principled harmonic concomitants of dimensional combination, we need ways of deriving Harmony scales from phonetic scales. Symbolically, we have

#### (94) **Harmony Scale from Interaction of Phonetic Scales**

$$\{a > b \dots\} \otimes \{x > y > \dots\} = ax > \dots$$

The goal of contextual markedness theory is to give content to the operator  $\otimes$ . Below in §8.2 we introduce a formal mechanism of *Prominence Alignment* which generates constraint rankings from paired phonetic scales, yielding a Harmony scale on their combination. In the syllable structure application of §8.2, the two phonetic scales which are aligned are segmental prominence (the

sonority dimension) and syllable position prominence (Peak is a more prominent position than Margin). The result is a Harmony scale on associations of segments to syllable positions.

It is important to distinguish the three kinds of scales or hierarchies which figure in Optimality Theory. To minimize confusions, we have given each its own distinctive comparison symbol. Two of these figure in (94): elements are ordered on a phonetic scale by the relation '>', and on a Harmony scale according to '>'. The third type of hierarchy in the theory is the domination hierarchy, along which constraints are ranked by the relation '>>'. These different types of scales are enumerated and exemplified in the following table:

(95) **Three Different Scales in Optimality Theory**

Type of scale or hierarchy	Relates	Symbol	Example	Meaning
Phonetic Scale	Points along elementary representational dimensions	>	$a > l$	$a$ is more sonorous than $l$
Harmony Scale	Well-formedness of structural configurations built from elementary dimensions	>	$\acute{a} > \acute{l}$	a nucleus filled by $a$ is more harmonic than a nucleus filled by $l$
Domination Hierarchy	Relative priority of well-formedness	>>	ONS>>HNUC	the constraint ONS strictly dominates the constraint HNUC

## 5.2 The Theory of Constraint Interaction

In order to define harmonic comparison of candidates consisting of entire parses, we will proceed in two steps. First, we get clear about comparing entire candidates on the basis of a single constraint, using ONS and HNUC from the Berber analysis in §2 as our examples. Then we show how to combine the evaluation of these constraints using a dominance hierarchy.

### 5.2.1 Comparison of Entire Candidates by a Single Constraint

The first order of business is a precise definition of how a single constraint ranks entire parses. We start with the simpler case of a single binary constraint, and then generalize the definition to non-binary constraints.



### 5.2.1.1 ONS: Binary constraints

It is useful to think of ONS as examining a syllable to see if it has an onset; if it does not, we think of ONS as assessing a **mark** of violation, \*ONS. ONS is an example of a *binary constraint*; a given syllable either satisfies or violates the constraint entirely. The marks ONS generates are all of the same type: \*ONS. For the moment, all the marks under consideration are identical. Later, when we consider the interaction of multiple binary constraints, there will be different types of marks to distinguish; each binary constraint  $C$  generates marks of its own characteristic type, \* $C$ . Furthermore, some constraints will be non-binary, and will generate marks of different types representing different degrees of violation of the constraint: the next constraint we examine, HNUC, will illustrate this.

When assessing the entire parse of a string, ONS examines each  $\sigma$  node in the parse and assesses one mark \*ONS for each such node which lacks an onset. Introducing a bit of useful notation, let  $A$  be a prosodic parse of an input string, and let  $\text{ONS}(A) = (*\text{ONS}, *C, \dots)$  be a list containing one mark \*ONS for each onsetless syllable in  $A$ . Thus for example  $\text{ONS}(.tx\acute{z}.i\acute{t}.) = (*\text{ONS})$ : the second, onsetless, syllable earns the parse  $.tx\acute{z}.i\acute{t}.$  its sole \*ONS mark. (Here we use  $\acute{z}$  to indicate that  $z$  is parsed as a nucleus.)

ONS provides a criterion for comparing the Harmony of two parses  $A$  and  $B$ ; we determine which of  $A$  or  $B$  is more harmonic ('less marked') by comparing  $\text{ONS}(A)$  and  $\text{ONS}(B)$  to see which contains fewer \*ONS marks. We can notate this as follows:

$$A \succ_{\text{ONS}}^{\text{parse}} B \text{ iff } \text{ONS}(A) \succ^{(*)} \text{ONS}(B)$$

where ' $\succ_{\text{ONS}}^{\text{parse}}$ ' denotes comparison of entire parses and ' $\text{ONS}(A) \succ^{(*)} \text{ONS}(B)$ ' means 'the list  $\text{ONS}(A)$  contains fewer marks \*ONS than the list  $\text{ONS}(B)$ '. (We will use the notation '(\*)' as a mnemonic for 'list of marks'.) If the lists are the same length, then we write<sup>45</sup>

$$A \approx_{\text{ONS}}^{\text{parse}} B \text{ iff } \text{ONS}(A) \approx^{(*)} \text{ONS}(B).$$

It is extremely important to realize that what is crucial to  $\succ^{(*)}$  is not numerical counting, but simply comparisons of more or less. This can be emphasized through a recursive definition of  $\succ^{(*)}$ , a definition which turns out to provide the basis for the entire Optimality Theory formalism for Harmony evaluation. The intuition behind this recursive definition is very simple.

---

<sup>45</sup> In §5.1 we define several formally distinct orders in terms of one another. At the risk of overburdening the notation, we use superscripts like <sup>parse</sup> and <sup>(\*)</sup> to keep all these orders distinct. We prefer to resist the temptation to sweep conceptual subtleties under the rug by using extremely concise notation in which many formally distinct relations are denoted by the same symbol. It is important to remember, however, that the symbols ' $\succ$ ' and ' $\approx$ ' — no matter what their subscripts and superscripts — always mean 'more harmonic' and 'equally harmonic'. We need to compare the Harmonies of many different kinds of elements, and for clarity while setting up the fundamental definitions of the theory, we distinguish these different Harmony comparison operators. Once the definitions are grasped, however, there is no risk of confusion in dropping superscripts and subscripts, which we will indeed do. The superscripts and subscripts can always be inferred from context — once the whole system is understood.

Suppose we are given two lists of identical marks  $*C$ ; we need to determine which list is shorter, and we can't count. Here's what we do. First, we check to see if either list is empty. If both are, the conclusion is that neither list is shorter. If one list is empty and the other isn't, the empty one is shorter. If neither is empty, then we remove one mark  $*C$  from each list, and start all over. The process will eventually terminate with a correct conclusion about which list is the shorter — but with no information about the numerical lengths of the lists.

Formalizing this recursive definition is straightforward; it is also worthwhile, since the definition will be needed anyway to characterize the full means of evaluating the relative harmonies of two candidate parses.

We assume two simple operations for manipulating lists. The operation we'll call **FM** extracts the First Member (or ForeMost element) of a list; this is what we use to extract the First Mark  $*C$  from each list. The other operation **Rest** takes a list, throws away its First Member, and returns the rest of the list; we use this for the recursive step of 'starting over', asking which list is shorter after the first  $*C$  has been thrown out of each.

Since we keep throwing out marks until none are left, it's also important to deal with the case of empty lists. We let  $()$  denote an empty list, and we define FM so that when it operates on  $()$ , its value is  $\emptyset$ , the null element.

Now let  $\alpha$  and  $\beta$  be two lists of marks. We write  $\alpha \succ^{(*)} \beta$  for ' $\alpha$  is more harmonic than  $\beta$ ', which in the current context means ' $\alpha$  is *shorter* than  $\beta$ ', since marks are anti-harmonic. To express the fact that an empty list of marks is more harmonic than a non-empty list, or equivalently that a null first element indicates a more harmonic list than does a non-null first element  $*C$ , we adopt the following relation between single marks:

(96) **Marks are anti-harmonic:**  $\emptyset \succ^{(*)} *C$

Remembering that  $\approx$  denotes 'equally harmonic' (or 'equally marked'), we also note the obvious facts about identical single marks:

$$\emptyset \approx^{(*)} \emptyset \text{ and } *C \approx^{(*)} *C$$

Our recursive definition of  $\succ^{(*)}$  can now be given as follows, where  $\alpha$  and  $\beta$  denote two lists of identical marks:

(97) **Harmonic ordering — lists of identical marks**

$\alpha \succ^{(*)} \beta$  iff either:

(i)  $FM(\alpha) \succ^{(*)} FM(\beta)$

or

(ii)  $FM(\alpha) \approx^{(*)} FM(\beta)$  and  $Rest(\alpha) \succ^{(*)} Rest(\beta)$

' $\beta \prec^{(*)} \alpha$ ' is equivalent to ' $\alpha \succ^{(*)} \beta$ '; ' $\alpha \approx^{(*)} \beta$ ' is equivalent to 'neither  $\alpha \succ^{(*)} \beta$  nor  $\beta \succ^{(*)} \alpha$ '.

(In subsequent order definitions, we will omit the obvious counterparts of the final sentence defining  $\prec^{(*)}$  and  $\approx^{(*)}$  in terms of  $\succ^{(*)}$ .)

To repeat the basic idea of the definition one more time in English:  $\alpha$  is shorter than  $\beta$  iff (if and only if) one of the following is true: (i) the first member of  $\alpha$  is null and the first member of  $\beta$  is not (*i.e.*,  $\alpha$  is empty and  $\beta$  is not), or (ii) the list left over after removing the first member of  $\alpha$  is shorter than the list left over after removing the first member of  $\beta$ .<sup>46</sup>

Now we can say precisely how ONS assesses the relative Harmony of two candidate parses, say  $.t\acute{x}.z\acute{i}t.$  and  $.t\acute{x}\acute{z}.i\acute{t}.$  ONS assesses the first as more harmonic than the second, because the second has an onsetless syllable and the first does not. We write this as follows:

$$.t\acute{x}.z\acute{i}t. \succ_{\text{ONS}}^{\text{parse}} .t\acute{x}\acute{z}.i\acute{t}. \text{ because } \text{ONS}(.t\acute{x}.z\acute{i}t.) = () \succ^{(*)} (*\text{ONS}) = \text{ONS}(.t\acute{x}\acute{z}.i\acute{t}.)$$

where  $\succ^{(*)}$  is defined in (97).

As another example:

$$.t\acute{x}.z\acute{i}t. \approx_{\text{ONS}}^{\text{parse}} .t\acute{x}\acute{z}.n\acute{i}. \text{ because } \text{ONS}(.t\acute{x}.z\acute{i}t.) = () \approx^{(*)} () = \text{ONS}(.t\acute{x}\acute{z}.n\acute{i}.)$$

In general, for any binary constraint  $\mathbb{C}$ , the harmonic ordering of entire parses which it determines,  $\succ_{\mathbb{C}}^{\text{parse}}$ , is defined as follows, where A and B are candidate parses:

**(98) Harmonic ordering of forms — entire parses, single constraint  $\mathbb{C}$**

$$A \succ_{\mathbb{C}}^{\text{parse}} B \text{ iff } \mathbb{C}(A) \succ^{(*)} \mathbb{C}(B)$$

with  $\succ^{(*)}$  as defined in (97).

It turns out that these definitions of  $\succ^{(*)}$  (97) and  $\succ_{\mathbb{C}}^{\text{parse}}$  (98), which we have developed for binary constraints (like ONS) apply equally to non-binary constraints (like HNUC); in the general case, a constraint's definition includes a harmonic ordering of the various types of marks it generates. The importance of the definition justifies bringing it all together in self-contained form:

<sup>46</sup> A simple example of how this definition (97) works is the following demonstration that

$$(*\mathbb{C}) \succ^{(*)} (*\mathbb{C}, *\mathbb{C}).$$

Define  $\alpha$  and  $\beta$  as follows (we use '=' for 'is defined to be'):

$$\alpha \equiv (*\mathbb{C}) \quad \beta \equiv (*\mathbb{C}, *\mathbb{C}).$$

Then

$$\alpha \succ^{(*)} \beta \text{ because}$$

$$(97.\text{ii}) \text{ FM}(\alpha) = *\mathbb{C} \approx^* *\mathbb{C} = \text{FM}(\beta) \text{ and}$$

$$\text{Rest}(\alpha) = () \succ^{(*)} (*\mathbb{C}) = \text{Rest}(\beta);$$

where the last line,  $() \succ^{(*)} (*\mathbb{C})$ , is in turn demonstrated by letting

$$\alpha' \equiv () \quad \beta' \equiv (*\mathbb{C})$$

and noting that

$$\alpha' \succ^{(*)} \beta' \text{ because}$$

$$(97.\text{i}) \text{ FM}(\alpha') = \emptyset \succ^* *\mathbb{C} = \text{FM}(\beta')$$

by (96).

(99)

<b>Harmonic ordering of forms — entire parse, single constraint</b>
<p>Let <math>\mathbb{C}</math> denote a constraint. Let A,B be two candidate parses, and let <math>\alpha, \beta</math> be the lists of marks assigned them by <math>\mathbb{C}</math>:</p> $\alpha \equiv \mathbb{C}(A), \beta \equiv \mathbb{C}(B)$ <p><math>\mathbb{C}</math> by definition provides a Harmony order <math>\succ^*</math> of the marks it generates. This order is extended to a Harmony order <math>\succ^{(*)}</math> over lists of marks as follows:</p> <p><math>\alpha \succ^{(*)} \beta</math> iff either:</p> <p style="padding-left: 40px;">(i) <math>\text{FM}(\alpha) \succ^* \text{FM}(\beta)</math></p> <p style="padding-left: 40px;">or</p> <p style="padding-left: 40px;">(ii) <math>\text{FM}(\alpha) \approx^* \text{FM}(\beta)</math> and <math>\text{Rest}(\alpha) \succ^{(*)} \text{Rest}(\beta)</math></p> <p>This order <math>\succ^{(*)}</math> is in turn extended to a Harmony order over candidate parses (with respect to <math>\mathbb{C}</math>), <math>\succ_{\mathbb{C}}^{\text{parse}}</math>, as follows:</p> $A \succ_{\mathbb{C}}^{\text{parse}} B \text{ iff } \mathbb{C}(A) \equiv \alpha \succ^{(*)} \beta \equiv \mathbb{C}(B)$

The case we have so far considered, when  $\mathbb{C}$  is binary, is the simplest precisely because the Harmony order over marks which gets the whole definition going,  $\succ^*$ , is so trivial:

$$\emptyset \succ^* * \mathbb{C}$$

‘a mark absent is more harmonic than one present’ (96). In the case we consider next, however, the ordering of the marks provided by  $\mathbb{C}$ ,  $\succ^*$ , is more interesting.

### 5.2.1.2 HNUC: Non-binary constraints

Turn now to HNUC. When it examines a single syllable, HNUC can usefully be thought of as generating a symbol designating the nucleus of that syllable; if the nucleus is  $n$ , then HNUC generates  $\acute{n}$ . HNUC arranges these nucleus symbols in a Harmony order, in which  $\acute{x} \succ_{\text{HNUC}} \acute{y}$  if and only if  $x$  is more sonorous than  $y$ :  $|x| > |y|$ .

If A is an entire prosodic parse, HNUC generates a list of all the nuclei in A. For reasons soon to be apparent, it will be convenient to think of HNUC as generating a list of nuclei *sorted from most to least harmonic, according to HNUC* — i.e., from most to least sonorous. So, for example,  $\text{HNUC}(.tx\acute{z}.\acute{it}.) = (\acute{z}, \acute{z})$ .

When HNUC evaluates the relative harmonies of two entire syllabifications A and B, it first compares the most harmonic nucleus of A with the most harmonic nucleus of B: if that of A is more sonorous, then A is the winner without further ado. Since the lists of nuclei  $\text{HNUC}(A)$  and  $\text{HNUC}(B)$  are assumed sorted from most to least harmonic, this process is simply to compare the First Member of  $\text{HNUC}(A)$  with the First Member of  $\text{HNUC}(B)$ : if one is more harmonic than the other, according to HNUC, the more harmonic nucleus wins the competition for its entire parse. If, on the other hand, the two First Members of  $\text{HNUC}(A)$  and  $\text{HNUC}(B)$  are equally harmonic according to HNUC (i.e., equally sonorous), then we eject these two First Members from their respective lists and start over, comparing the Rest of the nuclei in exactly the same fashion.

This procedure is exactly the one formalized above in (99). We illustrate the formal definition by examining how HNUC determines the relative harmonies of

$$A \equiv .t\acute{x}.z\acute{t}\acute{t}. \quad \text{and} \quad B \equiv .t\acute{x}.z\acute{t}\acute{t}.$$

First,  $\mathbb{C} \equiv \text{HNUC}$  assigns the following:

$$\alpha \equiv \mathbb{C}(A) = (\acute{t}, x) \quad \beta \equiv \mathbb{C}(B) = (\acute{t}, t)$$

To rank the parses A and B, *i.e.*, to determine whether

$$A \succ_{\mathbb{C}}^{\text{parse}} B,$$

we must rank their list of marks according to  $\mathbb{C}$ , *i.e.*, determine whether

$$\mathbb{C}(A) \equiv \alpha \succ^{(*)} \beta \equiv \mathbb{C}(B).$$

To do this, we examine the First Marks of each list, and determine whether

$$\text{FM}(\alpha) \succ^* \text{FM}(\beta).$$

As it happens,

$$\text{FM}(\alpha) \approx \text{FM}(\beta),$$

since both First Marks are  $\acute{t}$ , so we must discard the First Marks and examine the Rest, to determine whether

$$\alpha' \equiv \text{Rest}(\alpha) \succ^{(*)} \text{Rest}(\beta) \equiv \beta'.$$

Here,

$$\alpha' = (x); \quad \beta' = (t).$$

So again we consider First Marks, to determine whether

$$\text{FM}(\alpha') \succ^* \text{FM}(\beta').$$

Indeed this is the case:

$$\text{FM}(\alpha') = x \succ^* t = \text{FM}(\beta')$$

since  $|x| > |t|$ . Thus we finally conclude that

$$.t\acute{x}.z\acute{t}\acute{t}. \succ_{\text{HNUC}}^{\text{parse}} .t\acute{x}.z\acute{t}\acute{t}.$$

HNUC assesses nuclei  $\acute{x}$  from most to least harmonic, and that is how they are ordered in the lists HNUC generates for Harmony evaluation. HNUC is an unusual constraint in this regard; the other non-binary constraints we consider in this book will compare their *worst* marks first; the mark lists they generate are ordered from least- to most-harmonic. Both kinds of constraints are treated by the same definition (99). The issue of whether mark lists should be generated worst- or best-first will often not arise, for one of two reasons. First, if a constraint  $\mathbb{C}$  is binary, the question is meaningless because all the marks it generates are identical:  $*\mathbb{C}$ . Alternatively, if a constraint applies only once to an entire parse, then it will generate only one mark per candidate, and the issue of ordering multiple marks does not even arise. (Several examples of such constraints, including edgemostness of main stress, or edgemostness of an infix, are discussed in §4.) But for constraints like HNUC which are non-binary and which apply multiply in a candidate parse, part of the definition of the constraint must be whether it lists worst- or best-marks first.

### 5.2.2 Comparison of Entire Candidates by an Entire Constraint Hierarchy

We have now defined how a single constraint evaluates the relative Harmonies of entire candidate parses (99). It remains to show how a collection of such constraints, arranged in a strict domination hierarchy  $[\mathbb{C}_1 \gg \mathbb{C}_2 \gg \dots]$ , *together* perform such an evaluation: that is, how constraints interact.

Consider the part of the Berber constraint hierarchy we have so far developed: [ONS >> HNUC]. The entire hierarchy can be regarded as assigning to a complete parse such as *.txʒ.ít.* the following list of lists of marks:

$$(100.a) \quad [ONS \gg HNUC](.txʒ.ít.) = [ONS(.txʒ.ít.), HNUC(.txʒ.ít.)] = [(*ONS), (í, ʒ)]$$

The First Member here is the list of marks assigned by the dominant constraint: (\*ONS). Following are the lists produced by successive constraints down the domination hierarchy; in this case, there is just the one other list assigned by HNUC. As always, the nuclei are ordered from most- to least-harmonic by HNUC.

We use square brackets to delimit this list of lists, but this is only to aid the eye, and to suggest the connection with constraint hierarchies, which we also enclose in square brackets. Square and round brackets are formally equivalent here, in the sense that they are treated identically by the list-manipulating operations FM and Rest.

The general definition of the list of lists of marks assigned by a constraint hierarchy is simply:

#### (101) Marks Assigned by an Entire Constraint Hierarchy

The marks assigned to an entire parse A by a constraint hierarchy [C1 >> C2 >> ...] is the following list of lists of marks:

$$[C1 \gg C2 \gg \dots](A) = [C1(A), C2(A), \dots]$$

Consider a second example, *.tx.zít.*:

$$(100.b) \quad [ONS \gg HNUC](.tx.zít.) = [ONS(.tx.zít.), HNUC(.tx.zít.)] = [(), (í, x)]$$


Since there are no onsetless syllables in this parse,  $ONS(.tx.zít.) = ()$ , the empty list. A third example is:

$$(100.c) \quad [ONS \gg HNUC](.tx.zít.) = [(), (í, t)]$$

As always in Berber, the ONS constraint is lifted phrase-initially, so this parse incurs no marks \*ONS.

Now we are ready to harmonically rank these three parses. Corresponding directly to the example tableau (17) of §2, p.20, repeated here:

#### (102) Constraint Tableau for three parses of /txznt/

Candidates	ONS	HNUC
 <i>.tx.zít.</i>		ń x
<i>´tx.zít.</i>		ń t !
<i>.txʒ.ít.</i>	* !	ń ʒ

we have, from (100.a–c):

(103) **Marks Assessed by the Constraint Hierarchy** on three parses of /txznt/

A	[ONS >> HNUC] (A)
.tʰ.zní̄t.	[ ( ) , (ń ʰ) ]
.íx.zní̄t.	[ ( ) , (ń í) ]
.txʒ.nít.	[ (*ONS) , (ń ʒ) ]

To see how to define the Harmony order  $\succ_{[ONS \gg HNUC]}$  that the constraint hierarchy imposes on the candidate parses, let's first review how Harmony comparisons are performed with the tableau (102). We start by examining the marks in the first, ONS, column. Only the candidates which fare best by these marks survive for further consideration. In this case, one candidate, *.txʒ.nít.*, is ruled out because it has a mark \*ONS while the other two do not. That is, this candidate is less harmonic than the other two with respect to the hierarchy [ONS >> HNUC] because it is less harmonic than the other two with respect to the dominant individual constraint ONS. The remaining two parses *.tʰ.zní̄t.* and *.íx.zní̄t.* are equally harmonic with respect to ONS, and so to determine their relative Harmonies with respect to [ONS >> HNUC] we must continue by comparing them with respect to the next constraint down the hierarchy, HNUC. These two parses are compared by the individual constraint HNUC in just the way we have already defined: the most harmonic nuclei are compared first, and since this fails to determine a winner, the next-most harmonic nuclei are compared, yielding the final determination that *.tʰ.zní̄t.*  $\succ_{[ONS \gg HNUC]}$  *.íx.zní̄t.*

For the case of [ONS >> HNUC], the definition should now be clear:

(104) **Harmonic ordering of forms — entire parses by [ONS >> HNUC].**

- A  $\succ_{[ONS \gg HNUC]}$  B iff either
- (i) A  $\succ_{ONS}$  B
- or
- (ii) A  $\approx_{ONS}$  B and A  $\succ_{HNUC}$  B

For a general constraint hierarchy, we have the following recursive definition:

(105)

Harmonic ordering of forms — entire parse, entire constraint hierarchy
<p>A <math>\succ_{[C1 \gg C2 \gg \dots]}</math> B iff either</p> <ul style="list-style-type: none"> <li>(i) A <math>\succ_{C1}</math> B</li> </ul> <p>or</p> <ul style="list-style-type: none"> <li>(ii) A <math>\approx_{C1}</math> B and A <math>\succ_{[C2 \gg \dots]}</math> B</li> </ul>

All the orderings in (104) and (105) are of complete parses, and we have therefore omitted the superscript <sup>parse</sup>. The Harmony order presupposed by this definition,  $\succ_{C_1}^{\text{parse}}$ , the order on entire parses determined by the single constraint  $C_1$ , is defined in (99).

It is worth showing that the definitions of whole-parse Harmony orderings by a single constraint  $\succ_C$  (99) and by a constraint hierarchy  $\succ_{[C_1 \gg C_2 \gg \dots]}$  (105) are essentially identical. To see this, we need only bring in FM and Rest explicitly, and insert them into (105); the result is the following:

**(106) Harmonic ordering of forms — entire parses, entire constraint hierarchy** (opaque version).

Let  $\text{CH} \equiv [C_1 \gg C_2 \gg \dots]$  be a constraint hierarchy and let  $A, B$  be two candidate parses. Let  $\mathfrak{N}, \mathfrak{J}$  be the two lists of lists of marks assigned to these parses by the hierarchy:

$$\mathfrak{N} \equiv \text{CH}(A), \quad \mathfrak{J} \equiv \text{CH}(B)$$

It follows that:

$$\begin{aligned} \text{FM}(\mathfrak{N}) &= C_1(A), & \text{Rest}(\mathfrak{N}) &= [C_2 \gg \dots](A); \\ \text{FM}(\mathfrak{J}) &= C_1(B), & \text{Rest}(\mathfrak{J}) &= [C_2 \gg \dots](B) \end{aligned}$$

The hierarchy  $\text{CH}$  determines a harmonic ordering over lists of lists of marks as follows:

$\mathfrak{N} \succ^{[*]} \mathfrak{J}$  iff either

$$(i) \quad \text{FM}(\mathfrak{N}) \succ^{(*)} \text{FM}(\mathfrak{J}) \quad (\text{i.e., } C_1(A) \succ^{(*)} C_1(B), \text{ i.e., } A \succ_{C_1} B)$$

or

$$(ii) \quad \text{FM}(\mathfrak{N}) \approx^{(*)} \text{FM}(\mathfrak{J}) \quad (\text{i.e., } A \approx_{C_1} B)$$

and

$$\text{Rest}(\mathfrak{N}) \succ^{[*]} \text{Rest}(\mathfrak{J})$$

The harmonic ordering over candidate parses determined by  $\text{CH}$  is then defined by:

$$A \succ_{\text{CH}}^{\text{parse}} B \quad \text{iff} \quad \text{CH}(A) \equiv \mathfrak{N} \succ^{[*]} \mathfrak{J} \equiv \text{CH}(B)$$

This definition of  $\succ_{\text{CH}}^{\text{parse}}$  is identical to the definition of  $\succ_C^{\text{parse}}$  (99) except for the inevitable substitutions: the single constraint  $C$  of (99) has been replaced with a constraint hierarchy  $\text{CH}$  in (106), and, accordingly, one additional level has been added to the collections of marks.

The conclusion, then, is that whole-parse Harmony ordering by constraint hierarchies is defined just like whole-parse Harmony ordering by individual constraints. To compare parses, we compare the marks assigned them by the constraint hierarchy. This we do by first examining the First Marks — those assigned by the dominant constraint. If this fails to decide the matter, we discard the First Marks, take the Rest of the marks (those assigned by the remaining constraints in the hierarchy) and start over with them.

Thus, there is really only one definition for harmonic ordering in Optimality Theory; we can take it to be (99). The case of binary marks (§5.2.1.1) is a simple special case, where ‘less marked’ reduces to ‘fewer (identical) marks’; the case of constraint hierarchies (106) is a mechanical generalization gotten by making obvious substitutions.



### 5.2.3 Discussion

#### 5.2.3.1 Non-locality of interaction

As mentioned at the end of §2, the way that constraints interact to determine the Harmony ordering of an entire parse is somewhat counter-intuitive. In the Berber hierarchy [ONS >> HNUC], for example, perhaps the most obvious way of ordering two parses is to compare the parses syllable-by-syllable, assessing each syllable independently first on whether it meets ONS, and then on how well it fares with HNUC. As it happens, this can be made to work for the special case of [ONS >> HNUC] if we evaluate syllables in the correct order: from most- to least-harmonic. This procedure can be shown to more-or-less determine the same optimal parses as the different harmonic ordering procedure we have defined above, but only because some very special conditions obtain: first, there are only two constraints, and second, the dominant one is never violated in optimal parses.<sup>47</sup> Failing

---

<sup>47</sup> The argument goes as follows. Suppose A is the optimal parse of an input I according to harmonic ordering, so that  $A \succ B$  for any other parse B of I. We need to show that A beats B in the following syllable-by-syllable comparison: compare the most harmonic syllable in A with the most harmonic syllable in B; if one is more harmonic than the other, its parse wins; if they are equally harmonic, discard these two best syllables and recursively evaluate the remaining ones. To determine which of two syllables is the more harmonic, evaluate them by [ONS >> HNUC]; that is, compare them against each other on ONS and if that fails to pick a winner, compare them on HNUC.

Here's the argument. Every input has parses which do not violate the dominant constraint ONS, so parses judged optimal by harmonic ordering never violate ONS. Since A is optimal, none of its syllables violates ONS. Thus, in comparing any syllable  $\sigma_A$  of A with a syllable  $\sigma_B$  of B, A will win if  $\sigma_B$  violates ONS. So the only competitors B which could possibly beat A are those with no onsetless syllables. But in that case, ONS is irrelevant to the comparison of A and B: the syllable-by-syllable comparison will compare syllables from most- to least-harmonic based solely on HNUC. But this is exactly how the comparison goes according to harmonic ordering when both candidates have no violations of ONS. So the two methods must give the same result.

In short: ONS in either case serves to knock out all parses which violate it at all, and of the remaining parses the same one is picked out as optimal by the two methods because they both degenerate to the single remaining constraint HNUC.

This argument is correct regarding the core of the matter, but fails on a subtle issue: comparisons of parses with different numbers of syllables. In this case it can happen that the comparison has not yet been settled when one parse runs out of syllables and the other still has some remaining (which may or may not violate ONS). A definite procedure is required to handle the comparison of the null syllable  $\emptyset$  and a non-null syllable  $\sigma$ . And indeed here syllable-by-syllable comparison (but not harmonic ordering) fails: neither  $\emptyset \succ \sigma$  nor  $\sigma \succ \emptyset$  will work. To minimize distractions, consider the two hypothetical Berber inputs /tat/ and /tnmn/. The Dell-Elmedlaoui algorithm (and harmonic ordering) determine the corresponding outputs to be .*tát.* and .*tí.mí.* Now in order that .*tát.* beat .*tá.í.* in the syllable-by-syllable comparison, we must assume  $\emptyset \succ \sigma$ , because after the two best syllables (.*tát.* and .*tá.*) tie, the correct parse has no more syllables while the incorrect parse has the remaining (miserable) syllable .*í.* On the other hand, in order that the correct parse .*tí.mí.* beat .*tnmn.*, we must assume  $\sigma \succ \emptyset$ ; for now, after the two best syllables (say .*tí* [which  $\approx$  .*mí*], and .*tnmn.*) tie, the competitor has no more syllables while the correct parse still has one left (.*mí*).

(continued...)

such special conditions, however, harmonic ordering as defined above and as used in the remainder of this book gives results which, as far as we know, cannot be duplicated or even approximated using the more obvious scheme of syllable-by-syllable evaluation. Indeed, when we extend our analysis of Berber even one step beyond the simple pair of constraints ONS and HNUC (see §8), harmonic ordering clearly becomes required to get the correct results.<sup>48</sup>

---

<sup>47</sup> (...continued)

The intuition evoked here is that really  $\emptyset$  is better than a syllable which violates ONS but worse than one which satisfies ONS. What this intuition really amounts to, we claim, is that ONS-violating syllables should lose the competition for their parses, and that this should have priority over trying to maximize nuclear Harmony — exactly as formalized in harmonic ordering. It is precisely because the ONS-violating syllable *.t.* in *.tá.t.* is so *bad* that it ends up being considered too late to clearly lose against a proper syllable in the correct parse. That is, postponing consideration of ONS-violating syllables until after considering better ONS-satisfying ones is exactly backwards — yet this is what the syllable-by-syllable method requires. Our diagnosis is that correctly handling ONS requires considering first those syllables that violate it — the *worst* syllables, according to ONS; whereas correctly handling HNUC requires considering first the syllables it rates the *best*. This is just what harmonic ordering does, by virtue of having separate passes over the parse for ONS and later for HNUC. Syllable-by-syllable evaluation is fundamentally incorrect in forcing one pass through the parse, evaluating each syllable according to both constraints and then discarding it. It is fortuitous that in the case of [ONS >> HNUC], in the majority of cases, best-first syllable comparison gives the correct answer: as long as a competitor has the same number of syllables as the optimal parse, an ONS violation will eventually get caught — even though such violations should really be handled first rather than last. Such postponement is revealed for the mistake it really is when parses with different numbers of syllables are examined.

<sup>48</sup> Our more complete analysis of Berber will include, among others, a universal constraint –COD which states that syllables must not have codas (this constraint is introduced in §4.1). This constraint is lower-ranked in the Berber hierarchy than HNUC, so a slightly more complete Berber hierarchy is [ONS >> HNUC >> –COD]. Now consider the input /*ratlult*/; the Dell-Elmedlaoui algorithm (and harmonic ordering) parses this as *.rát.lú.lf.* (the final stop desyllabifies in a subsequent process which as promised we ignore here). Note that the most harmonic syllable in this correct parse, *.rát.*, has a coda, violating –COD. Note further that there is a competing parse without the coda (which also respects ONS): *.rá.tlú.f.* Since HNUC >> –COD, the most harmonic syllable in this competing parse is also the one with the most sonorous nucleus, *.rá.* Now if we compare the most harmonic syllables of these two parses, we see that the correct parse *loses* because of its violation of –COD. (In case it is unclear whether comparing *least* harmonic syllables first might work, note that this also gives the wrong result here, since of the two parses being compared here, the correct parse contains the worst syllable: *.lf.*)

Again, harmonic ordering gets the correct result in comparing these two parses. First, ONS is checked throughout the parses; both respect it so the next constraint is considered. HNUC now declares the correct parse the winner, since its nuclei are more harmonic than those of its competitor: (*á, ú, t*) > (*á, l, f*). The competition is correctly resolved without ever consulting the lowest constraint –COD, which in this case can only lead the evaluation astray.

This example illustrates a kind of non-local constraint interaction captured by harmonic ordering but missed in the syllable-by-syllable approach. In order for the second syllable of the correct parse *.rát.lú.lf.* to optimize its nucleus (*ú*), the first syllable pays with a coda. (The parse *.rá.tlú.lf.* is blocked by a high-ranking  
(continued...)

It is important to note also that harmonic ordering completely finesses a nasty conceptual problem which faces a syllable-by-syllable approach as soon as we expand our horizons even slightly. For in general we need to rank complex parses which contain much more structure than mere syllables. The ‘syllable-by-syllable’ approach is conceptually really a ‘constituent-by-constituent’ approach, and in the general case there are many kinds and levels of constituents in the parse. Harmonic ordering completely avoids the need to decide in the general case how to correctly break structures into parts for Harmony evaluation so that, in part-by-part evaluation, all the relevant constraints have the proper domains for their evaluation. In harmonic ordering, each constraint  $\mathbb{C}$  independently generates its own list of marks  $\mathbb{C}(A)$  for evaluating a parse  $A$ , considering whatever domains within  $A$  are appropriate to that constraint. In comparing  $A$  with parse  $B$ , the marks  $\mathbb{C}(A)$  are compared with the marks  $\mathbb{C}(B)$ ; implicitly, this amounts to comparing  $A$  and  $B$  with respect to the domain structure peculiar to  $\mathbb{C}$ . This comparison is decoupled from that based on other constraints which may have quite different domain structure.

The interaction of constraints in a constituent-by-constituent approach is in a sense limited to interactions within a constituent: for ultimately the comparison of competing parses rests on the assessment of the Harmony of individual constituents as evaluated by the set of constraints. Optimality Theory is not limited to constraint interactions which are local in this sense, as a number of the subsequent analyses will illustrate (see also fn. 48).

### 5.2.3.2 Strictness of domination

Our expository example [ONS  $\gg$  HNUC] in Berber may fail to convey just how strong a theory of constraint interaction is embodied in harmonic ordering. In determining the correct — optimal — parse of an input, as the constraint hierarchy is descended, each constraint acts to disqualify remaining competitors with absolute independence from all other constraints. A parse found wanting on one constraint has absolutely no hope of redeeming itself by faring well on any or even all lower-ranking constraints. It is remarkable that such an extremely severe theory of constraint interaction has the descriptive power it turns out to possess.

Such strict domination of constraints is less striking in the Berber example we have considered than it will be in most subsequent examples. This is because the dominant constraint is never violated in the forms of the language; it is hardly surprising then that it has strict veto power over the lower constraint. In the general case, however, most of the constraints in the hierarchy will *not* be unviolated like ONS is in Berber. Nonetheless, *all* constraints in Optimality Theory, whether violated or not in the forms of the language, have the same strict veto power over lower constraints that ONS has in Berber.

---

<sup>48</sup> (...continued)

constraint which limits onsets to one segment, except phrase-initially; this is part of our fuller account of Berber in §8.) Raising the Harmony of the second syllable w.r.t HNUC at the cost of lowering the Harmony of the first w.r.t. –COD is in this case optimal because HNUC  $\gg$  –COD. However, HNUC and –COD interact here *across syllables*; and, in fact, the best syllable must pay to improve the less-good syllable. The syllable-by-syllable theory cannot correctly handle this non-local interaction, as we have seen; it wrongly rules against the correct parse because its best syllable has sacrificed Harmony (on a low-ranked constraint), never getting to the next-best syllable to see that it has improved Harmony (on a higher-ranked constraint).

### 5.2.3.3 Serial vs. Parallel Harmony Evaluation and Gen

Universal grammar must also provide a function Gen that admits the candidates to be evaluated. In the discussion in §2 we have entertained two different conceptions of Gen. The first, closer to standard generative theory, is based on serial or derivational processing; some general procedure (Do- $\alpha$ ) is allowed to make a certain single modification to the input, producing the candidate set of all possible outcomes of such modification. This is then evaluated; and the process continues with the output so determined. In this serial version of grammar, the theory of rules is narrowly circumscribed, but it is inaccurate to think of it as trivial. There are constraints inherent in the limitation to a single operation; and in the requirement that each individual operation in the sequence improve Harmony. (An example that springs to mind is the Move-x theory of rhythmic adjustments in Prince 1983; it is argued for precisely on the basis of entailments that follow from these two conditions, pp. 31-43.)

In the second, parallel-processing conception of Gen, all possible ultimate outputs are contemplated at once. Here the theory of operations is indeed rendered trivial; all that matters is what structures are admitted. Much of the analysis given in this book will be in the parallel mode, and some of the results will absolutely require it. But it is important to keep in mind that the serial/parallel distinction pertains to Gen and not to the issue of harmonic evaluation per se. It is an empirical question of no little interest how Gen is to be construed, and one to which the answer will become clear only as the characteristics of harmonic evaluation emerge in the context of detailed, full-scale, depth-plumbing, scholarly, and responsible analyses.<sup>49</sup>

---

<sup>49</sup> A faithful reconstruction of the sequential parsing process of the Dell-Elmedlaoui algorithm seems to be possible within the harmonic serial approach. Here is a quick sketch; the analysis has not been well developed. Many ideas are imported which will later be developed in the text in the context of the parallel approach. It seems highly unlikely that the sequential approach presented here for Berber can be extended to a sequential account of syllabification more generally.

The process starts with an input. Gen then generates a set of alternatives which are ‘one change away’ from the input. We let Gen perform any one of the following ‘changes’:

- (a) build a new syllable from free segments;
- (b) adjoin a free element to an existing syllable;
- (c) mark a segment x as surface-free:  $\langle x \rangle$ .

A surface-free segment is not phonetically realized in the surface form, as though deleted by Stray Erasure. Once a segment has been marked as surface-free it no longer counts as ‘free’; it is no longer available to Gen for operations (a–c). Berber never chooses to exercise the surface-free option, but we will derive this as a result rather than assuming it. The syllables constructed in (a) may include empty syllable positions which denote epenthetic elements; again, an option we show Berber not to exercise.

Now initially all segments in the input are free. Gen produces a set of candidates each of which is generated from the input by applying one of the ‘changes’ (a–c). The most harmonic of these is chosen as the next representation in the derivation. At each step at least one segment which was free becomes no longer free. The process is repeated until no free elements remain; this is the output.

At each step of the derivation, the candidates generated by Gen each contain one ‘changed’ element; for (a), it is a newly constructed syllable; for (b), a syllable with a segment newly adjoined; for (c), a segment marked surface-free. To compare the Harmonies of two such candidates, we simply compare the one changed

(continued...)

<sup>49</sup> (...continued)

element of each candidate; the remaining parts of the two candidates are identical. If both candidates are generated by either (a) or (b), then we are comparing two syllables. This we can do using a constraint hierarchy, as previously explained in the text. The hierarchy we assume for Berber is as follows:

NUC >> \*COMPLEX >> PARSE >> FILL >> ONS >> -COD >> HNUC

This is virtually identical to the hierarchy of the parallel analysis we develop in §8.1.1 (ignoring exceptional epenthesis), except that -COD is higher-ranked in this sequential analysis, for reasons to be discussed. The constraints are more fully developed in the text; here is a quick summary. NUC requires nuclei; \*COMPLEX forbids more than one segment in onset, nucleus, or coda; PARSE forbids surface-free segments; FILL forbids empty (epenthetic) syllable positions; -COD forbids codas.

At each step Gen generates via (a) candidates each with a new syllable. When these are compared using the constraint hierarchy, the most harmonic such new syllables will always have a single segment onset and a single segment nucleus, with no epenthetic positions and no coda. That is, they will always be core syllables. Gen generates all sorts of new syllables, but the most harmonic will always be those built as  $xy \rightarrow \{xY\}$ , for only these satisfy all of the top six constraints. Of these, the most harmonic will be determined by the seventh constraint, HNUC. As long as there is a free pair of adjacent segments  $xy$ , a candidate generated from  $xy \rightarrow \{xY\}$  via (a) will be more harmonic than all candidates generated via adjunction (b) or surface-free marking (c); for adjunction to a core syllable already built (b) will violate \*COMPLEX unless the adjunction is to coda position, in which case it will violate -COD; and a surface-free marking (c) violates PARSE. No such violations occur with  $\{xY\}$ . Thus, as long as free pairs  $xy$  exist, the most harmonic candidate generated by Gen will always be the one which performs  $xy \rightarrow \{xY\}$  where Y is the most sonorous available such segment. This is of course exactly the principal step of the Dell-Elmedlaoui algorithm.

The final part of the Dell-Elmedlaoui algorithm takes place when there are no longer any free pairs  $xy$ . Then free singleton segments are adjoined as codas to the preceding already-built core syllable. This too is reconstructed by harmonic sequential parsing. For when there are no longer free pairs  $xy$ , any new syllables generated via (a) by Gen are no longer the most harmonic changes. Such a new syllable must be erected over a single underlying segment  $x$  (or be totally epenthetic), and this syllable must therefore violate at least one of the constraints NUC, FILL, or ONS. These constraints are all higher ranked than -COD, which is the only constraint violated by the changed element in candidates generated via (b) by adjoining  $x$  as the coda of an existing core syllable. The other candidates are generated via (c) by marking  $x$  as surface-free; the changed element in such candidates,  $\langle x \rangle$ , violates PARSE, which dominates -COD, so surface-free candidates (c) are less harmonic than coda-adjoined ones (b). At each step in this last phase of parsing, then, one free singleton segment will be coda-adjoined, until finally there are no free segments left and parsing is complete.

For this sequential approach to work, it is crucial that -COD >> HNUC — although in the parallel approach developed in the text, it is equally crucial that HNUC >> -COD. The sequential parsing algorithm must build all possible core syllables before creating any codas; the location of codas in the correct parse can only be determined after the nuclei have been taken in descending sonority order. -COD >> HNUC ensures that a newly-closed syllable will be sub-optimal as long as new core syllables exist in the candidate set. To see what would happen if HNUC >> -COD, consider the example input /ratlult/, considered in (6), p.14, §2.1. The most harmonic first step is  $\{\mathbf{ra}\}t\mathbf{lult}$ . Now the next step should be to  $\{\mathbf{ra}\}t\{\mathbf{lu}\}lt$ , with changed element  $\{\mathbf{lu}\}$ . But consider the adjunction (b) candidate  $\{\mathbf{rat}\}lult$ , with changed element  $\{\mathbf{rat}\}$ . On HNUC,  $\{\mathbf{rat}\}$  bests  $\{\mathbf{lu}\}$ , while on -COD,  $\{\mathbf{lu}\}$  is preferred. Thus the right choice will only be made if -COD >> HNUC.

To most clearly see why the parallel approach requires HNUC >> -COD, consider the hypothetical input /tat/, which the Dell-Elmedlaoui algorithm parses as  $\{\mathbf{tat}\}$ , a closed syllable. An alternative complete

(continued...)

Many different theories of the structure of phonological outputs can be equally well accommodated in Gen, and the framework of Optimality Theory *per se* involves no commitment to any set of such assumptions. Of course, different structural assumptions can suggest or force different formal approaches to the way that Optimality theoretic constraints work. In this work, to implement faithfulness straightforwardly, we entertain a non-obvious assumption about Gen which will be useful in implementing the parallel conception of the theory: we will assume, following the lead of McCarthy 1979 and Itô 1986, 1989, that every output for an input *In* – every member of  $\text{Gen}(In)$  – includes *In* as an identifiable substructure. In the theory of syllable structure developed in Part II,  $\text{Gen}(/txznt/)$  will be a set of possible syllabifications of */txznt/* all of which contain the input string */txznt/*, with each underlying segment either associated to syllable structure or left unassociated. We will interpret unassociated underlying segments as phonetically unrealized (*cf.* ‘Stray Erasure’); thus on this conception, input segments are never ‘deleted’ in the sense of disappearing from the structural description; rather, they may simply be left free — unparsed. Our discussion of Berber in this section has focused on a fairly restricted subset of the full candidate set we will subsequently consider; we have considered only syllabifications in which underlying segments are in one-to-one correspondence with syllable positions. In following chapters, we turn to languages which, unlike Berber, exhibit syllabifications manifesting deletion and/or epenthesis.

### 5.2.3.4 Binary vs. Non-binary constraints

As might be suspected, it will turn out that the work done by a single non-binary constraint like HNUC can also be done by a *set* (indeed a sub-hierarchy) of binary constraints. This will prove fundamental for the construction of the Basic Segmental Syllable Theory in §8, and we postpone treatment of the issue until then. For now it suffices simply to remark that the division of constraints into those which are binary and those which are not, a division which we have adopted earlier in this section, is not in fact as theoretically fundamental as it may at this point appear.

## 5.3 Pāṇini’s Theorem on Constraint Ranking

One consequence of the definition of harmonic ordering is that there are conditions under which the presence of a more general constraint in a superordinate position in a hierarchy will eliminate all opportunities for a more specialized constraint in a subordinate position to have any effects in the grammar. The theorem states, roughly, that if one constraint is more general than another in the sense that the set of inputs to which one constraint applies nonvacuously includes the other’s nonvacuous input set, and if the two constraints *conflict* on inputs to which the more specific applies nonvacuously, then the more specific constraint must dominate the more general one in order for its effects to be visible in the grammar. (This is an oversimplified first cut at the true result; such claims

---

<sup>49</sup> (...continued)

parse with only open syllables is  $\{T\}\{aT\}$ . In the parallel approach, we first scan the complete parse for violations of the top-ranked constraint; if this were –COD, then the correct parse would lose immediately, and its redeeming qualities with respect to lower-ranked HNUC would be irrelevant.

must be stated carefully.) Intuitively, the idea is that if the more specific constraint were lower-ranked, then for any input to which it applies non-vacuously, its effects would be over-ruled by the higher-ranked constraint with which it conflicts. The utility of the result is that it allows the analyst to spot certain easy ranking arguments.

We call this Pāṇini's Theorem on Constraint-ranking, in honor of the first known investigator in the area; in §7.2.1, we discuss some relations to the Elsewhere Condition of Anderson 1969 and Kiparsky 1973b. In this section we introduce some concepts necessary to develop a result; the proof is relegated to the Appendix. The result we state is undoubtedly but one of a family of related theorems which cover cases in which one constraint hides another.

Due to the complexities surrounding this issue, we will formally state and prove the result only in the case of constraints which are Boolean at the whole-parse level: constraints which assign a single mark to an entire parse when they are violated, and no mark when they are satisfied.

(107) **Dfn. Separation.** A constraint  $\mathbb{C}$  *separates* a set of structures if it is satisfied by some members of the set and violated by others.

(108) **Dfn. Non-vacuous application.** A constraint  $\mathbb{C}$  *applies non-vacuously* to an input  $i$  if it separates  $\text{Gen}(i)$ , the set of candidate parses of  $i$  admitted by Universal Grammar.

A constraint may sometimes apply vacuously to an input, in that every possible parse of  $i$  satisfies the constraint. For example, in §7 we will introduce a constraint  $\text{FREE-V}$  which requires that stem-final vowels *not* be parsed into syllable structure. Clearly, this constraint is vacuously satisfied for a stem which is not vowel-final; all the parses of such an input meet the constraint since none of them have a stem-final vowel which is parsed!

(109) **Dfn. Accepts.** A constraint hierarchy  $\mathbb{CH}$  *accepts* a parse  $P$  of an input  $i$  if  $P$  is an optimal parse of  $i$ .

When  $\mathbb{CH}$  is the entire constraint hierarchy of a grammar, it is normally the case that only one parse  $P$  of an input  $i$  is optimal: the constraint set is sufficient to winnow the candidate set down to a single output. In this section we will need to consider, more generally, initial portions of the constraint hierarchy of a grammar, *i.e.*, all the constraints from the highest-ranked down to some constraint which may not be the lowest-ranked. In these cases,  $\mathbb{CH}$  will often consist of just a few constraints, insufficient to winnow the candidate set down to a single parse; in that case,  $\mathbb{CH}$  will accept an entire set of parses, all equally harmonic, and all more harmonic than the competitors filtered out by  $\mathbb{CH}$ .

(110) **Dfn. Active.** Let  $\mathbb{C}$  be a constraint in a constraint hierarchy  $\mathbb{CH}$  and let  $i$  be an input.  $\mathbb{C}$  is *active on  $i$  in  $\mathbb{CH}$*  if  $\mathbb{C}$  separates the candidates in  $\text{Gen}(i)$  which are admitted by the portion of  $\mathbb{CH}$  which dominates  $\mathbb{C}$ .

In other words, the portion of  $\mathbb{CH}$  which dominates  $\mathbb{C}$  filters the set of candidate parses of  $i$  to some degree, and then  $\mathbb{C}$  filters it further. When  $\mathbb{C}$  is not active for an input  $i$  in  $\mathbb{CH}$ , the result of parsing  $i$  is not at all affected by the presence of  $\mathbb{C}$  in the hierarchy.

(111) Dfn. **Pāṇinian Constraint Relation.** Let  $\mathbb{S}$  and  $\mathbb{G}$  be two constraints.  $\mathbb{S}$  stands to  $\mathbb{G}$  as special to general in a Pāṇinian relation if, for any input  $i$  to which  $\mathbb{S}$  applies non-vacuously, any parse of  $i$  which satisfies  $\mathbb{S}$  fails  $\mathbb{G}$ .

For example, the constraint FREE-V stands to PARSE as special to general in a Pāṇinian relation: for any input to which FREE-V applies non-vacuously (that is, to any input with a stem-final vowel V), any parse which satisfies FREE-V (that is, which leaves V unparsed) must violate PARSE (in virtue of leaving V unparsed). For inputs to which the more specialized constraint FREE-V does *not* apply non-vacuously (C-final stems), the more general constraint PARSE need not conflict with the more specific one (for C-final stems, FREE-V is vacuously satisfied, but PARSE is violated in some parses and satisfied in others).

Now we are finally set to state the theorem:

(112) **Pāṇini's Theorem on Constraint-ranking.** Let  $\mathbb{S}$  and  $\mathbb{G}$  stand as specific to general in a Pāṇinian constraint relation. Suppose these constraints are part of a constraint hierarchy  $\mathbb{CH}$ , and that  $\mathbb{G}$  is active in  $\mathbb{CH}$  on some input  $i$ . Then if  $\mathbb{G} \gg \mathbb{S}$ ,  $\mathbb{S}$  is not active on  $i$ .

In §7, we will use this theorem to conclude that in the grammar of Lardil, the more specific constraint FREE-V must dominate the more general constraint PARSE with which it conflicts.