Optimality Theory in Phonology

A Reader | Edited by John J. McCarthy

Blackwell
Publishing

Optimality Theory in Phonology

Optimality Theory in Phonology

A Reader | Edited by John J. McCarthy

Blackwell
Publishing

350 Main Street, Malden, MA 02148-5020, USA
108 Cowley Road, Oxford OX4 1JF, UK
550 Swanston Street, Carlton, Victoria 3053, Australia

First published 2004 by Blackwell Publishing Ltd

Library of Congress Cataloging-in-Publication Data

Optimality theory in phonology : a reader / edited by John J. McCarthy.
p. cm.
Includes bibliographical references and index.
ISBN 0-631-22688-5 (hard. : alk. paper) — ISBN 0-631-22689-3 (pbk. : alk. paper)
1. Grammar, Comparative and general—Phonology. 2. Optimality theory (Linguistics) I. McCarthy, John J., 1953–
P217.3.067 2003
415—dc21
2003005153

A catalogue record for this title is available from the British Library.

Set in 10/12pt Sabon
by Graphicraft Limited, Hong Kong
Printed and bound in the United Kingdom
by T. J. International, Padstow, Cornwall

For further information on
Blackwell Publishing, visit our website:
http://www.blackwellpublishing.com

Contents

Preface

Optimality Theory (OT) has applications throughout the field of linguistics. But its first and greatest influence has been in phonology. This book is a compilation of readings on OT in phonology, starting with the original and most important one, Prince and Smolensky's *Optimality Theory: Constraint Interaction in Generative Grammar*. The readings cover a broad range of topics in phonology and related disciplines. Both previously published and never before published works are included.

The readings have been selected with a second-semester phonology course in mind, though they would also be suitable for a seminar or for independent reading. To enhance this work's usefulness as a textbook, I have included brief introductory notes at the beginning of each chapter to set the stage and point out connections with other chapters. Each chapter also concludes with a list of study and research questions. The questions appear in approximate order of difficulty: some are relatively easy reviews of the material; some are more challenging, requiring further thought and research; others are open-ended research topics and even notoriously unsolved problems, included here in the hope that they will elicit an answer. There is ample material for homework exercises, term papers, and dissertations in this book and in these study questions.

The decisions about what to include were extremely difficult, and many excellent works had to be omitted. When I was in doubt, considerations of length were decisive: articles that were already short or that could easily be made short were given priority. It is safe to say that no one else would make exactly these decisions, nor would I, I am sure, if I started all over again.

Almost without exception, the chapters of this book are excerpts from the original works. To cover a wide range of topics within limited space, I had to be severe in making cuts. If the original article had three sections, each describing a different example, two were cut. If there was interesting discussion that strayed from the main point, it was removed. Acknowledgments, digressive footnotes, appendices, and the like were excised automatically. Such minor omissions are not indicated in the text, though major ones are marked with "[. . .]", and the original numbering of sections, examples, and notes is retained, as are most of the cross-references. The excisions were all negotiated with the contributors, who gave their (sometimes reluctant) approval to the result. In some cases, authors went further, revising their chapters to smooth out the seams.

Each chapter includes the bibliography from the original work (minus any references that were cited only in the excised material). To these original bibliographies, two sigla have been added. The symbol ◀ marks references to works that are included in this reader. (The ◀ is mnemonic for "look at the table of contents in the front of the book.") The symbol ▶ marks references that are incomplete; it points to the bibliography at the back of the book, which includes better versions of those references plus all works cited in the Editor's Notes and Study and Research Questions. The notes and exercises also supplement the individual chapter bibliographies by pointing to more recent literature.

I would not have embarked on this project without instigation from Tami Kaplan and continuing support from her and from Sarah Coleman, both with Blackwell. I could not have completed it without the assistance of Maria Gouskova, whose care, common sense, and wisdom have been indispensable as the book came together. The index was largely the work of Michael Becker, Kathryn Flack, and Shigeto Kawahara, who did the job with remarkable care and swiftness. You readers and I owe a large debt to Margaret Aherne, copy-editor without peer, who not only turned a huge, messy manuscript into a handsome book, but also detected more than a few errors of substance in constraints, tableaux, and arguments. To all of these and to the authors, who have been generous with their time and help, I am very grateful.

John J. McCarthy
Amherst, Massachusetts
2003

Contributors' Addresses

John Alderete
Linguistics Program
406 Welsh Humanities Building
University of South Carolina
Columbia, SC 29208, USA

Arto Anttila
Department of Linguistics
New York University
719 Broadway, 5/F
New York, NY 10003, USA
arto.anttila@nyu.edu

Eric Baković
Linguistics Department
University of California, San Diego
9500 Gilman Drive, #0108
La Jolla, CA 92093, USA
bakovic@ling.ucsd.edu

Jill Beckman
Linguistics Department
570 English Philosophy Building
The University of Iowa
Iowa City, IA 52242, USA
jill-beckman@uiowa.edu

Laura Benua
40 Third Avenue
Nyack, NY 10960, USA
lbenua@hotmail.com

Paul de Lacy
Department of Linguistics
Faculty of Modern and Medieval Languages
University of Cambridge
Sidgwick Avenue
Cambridge CB3 9DA, England
pvd22@cam.ac.uk

Nine Elenbaas
Albert van Dalsumlaan 187
3584 HD Utrecht
The Netherlands
nine_elenbaas@hotmail.com

Bruce Hayes
Department of Linguistics
University of California
Los Angeles, CA 90095, USA
bhayes@humnet.ucla.edu

Sharon Inkelas
Department of Linguistics
1203 Dwinelle Hall
University of California
Berkeley, CA 94720, USA
inkelas@socrates.berkeley.edu

Junko Itô
Department of Linguistics
University of California, Santa Cruz
1156 High Street
Santa Cruz, CA 95064, USA
ito@ling.ucsc.edu

René Kager
Utrecht Institute of Linguistics OTS
Universiteit Utrecht
Trans 10
3512 JK Utrecht
The Netherlands
rene.kager@let.uu.nl

Michael Kenstowicz
Department of Linguistics & Philosophy, E39-245
MIT
Cambridge, MA 02139, USA
kenstow@mit.edu

Robert Kirchner
Department of Linguistics
4-32 Assiniboia Hall
University of Alberta
Edmonton, Alberta T6G 2E7, Canada
kirchner@ualberta.ca

Linda Lombardi
Linguistics Department
Marie Mount Hall 1401
University of Maryland
College Park, MD 20742, USA
linda_lombardi@umail.umd.edu

Anna Łubowicz
Department of Linguistics
3601 Watt Way
Grace Ford Salvatori Hall 301
University of Southern California
Los Angeles, CA 90089, USA

Joan Mascaró
Departament de Filologia Catalana
Universitat Autònoma de Barcelona
08193 Bellaterra, Spain
joan.mascaro@uab.es

John J. McCarthy
Department of Linguistics
University of Massachusetts
Amherst, MA 01003, USA
jmccarthy@linguist.umass.edu

Armin Mester
Department of Linguistics
University of California, Santa Cruz
1156 High Street
Santa Cruz, CA 95064, USA
mester@ling.ucsc.edu

Elliott Moreton
Department of Linguistics
CB #3155
Dey Hall 320
University of North Carolina
Chapel Hill, NC 27599, USA

Scott Myers
Department of Linguistics
University of Texas at Austin
Austin, TX 78712, USA
s.myers@mail.utexas.edu

Orhan Orgun
Linguistics Department
University of California, Davis
Davis, CA 95616, USA
ocorgun@ucdavis.edu

Jaye Padgett
Department of Linguistics
University of California, Santa Cruz
1156 High Street
Santa Cruz, CA 95064, USA
padgett@cats.ucsc.edu

Joe Pater
Department of Linguistics
University of Massachusetts
Amherst, MA 01003, USA
pater@linguist.umass.edu

Alan Prince
Department of Linguistics
18 Seminary Place
Rutgers University
New Brunswick, NJ 08903, USA
prince@ruccs.rutgers.edu

Jerzy Rubach
Fall
Linguistics Department
570 English Philosophy Building
The University of Iowa
Iowa City, IA 52242, USA
jerzy_rubach@uiowa.edu

Spring
Instytut Anglistyki
Uniwersytet Warszawski
ul. Nowy Swiat 4
00-497 Warszawa, Poland
rubach@mail.uw.edu.pl

Elisabeth Selkirk
Department of Linguistics
University of Massachusetts
Amherst, MA 01003, USA
selkirk@linguist.umass.edu

Paul Smolensky
Department of Cognitive Science
239A Krieger Hall
Johns Hopkins University
Baltimore, MD 21218, USA
smolensky@jhu.edu

Bruce Tesar
Department of Linguistics
18 Seminary Place
Rutgers University
New Brunswick, NJ 08903, USA
tesar@ruccs.rutgers.edu

Moira Yip
University College London
Department of Phonetics and Linguistics
Gower Street
London WC1E 6BT, England
moira@linguistics.ucl.ac.uk

Young-mee Yu Cho
Department of Asian Languages and Cultures
330 Scott Hall, College Avenue Campus
Rutgers University
New Brunswick, NJ 08901, USA
yucho@rci.rutgers.edu

Cheryl Zoll
Department of Linguistics & Philosophy,
E-39-245
MIT
Cambridge, MA 02139, USA
czoll@mit.edu

Acknowledgments

The editor and publisher wish to thank the following for permission to use copyright material in this book:

1 Prince, Alan and Smolensky, Paul (1993) *Optimality Theory: Constraint Interaction in Generative Grammar*. New Brunswick, NJ: Rutgers University Center for Cognitive Science. Technical report RuCCS-TR-2. Reprinted by permission of Alan Prince and Paul Smolensky.

2 McCarthy, John J. and Prince, Alan (1993) Generalized alignment. In *Yearbook of Morphology*, ed. Geert Booij and Jaap van Marle, pp. 79–153. Dordrecht: Kluwer. © 1993 by Kluwer Academic Publishers. Reprinted with kind permission of Kluwer Academic Publishers.

3 McCarthy, John J. and Prince, Alan (1999) Faithfulness and identity in Prosodic Morphology. In *The Prosody–Morphology Interface*, ed. René Kager, Harry van der Hulst, and Wim Zonneveld, pp. 218–309. Cambridge: Cambridge University Press. © 1999 by Cambridge University Press.

4 Tesar, Bruce (1995) *Computing Optimal Forms in Optimality Theory: Basic Syllabification*. Report no. CU-CS-763-95. Boulder, CO: Department of Computer Science, University of Colorado. Reprinted by permission of Bruce Tesar.

5 Tesar, Bruce and Smolensky, Paul (1998) Learnability in Optimality Theory. *Linguistic Inquiry* 29: 229–68. © 1998 by the Massachusetts Institute of Technology.

6 Moreton, Elliott (1996/1999) Non-computable Functions in Optimality Theory. Unpublished manuscript. Amherst, MA: University of Massachusetts, Amherst. Reprinted by permission of Elliott Moreton.

7 McCarthy, John J. and Prince, Alan (1993) Generalized alignment. In *Yearbook of Morphology*, ed. Geert Booij and Jaap van Marle, pp. 79–153. Dordrecht: Kluwer. © 1993 by Kluwer Academic Publishers. Reprinted with kind permission of Kluwer Academic Publishers.

8 Elenbaas, Nine and Kager, René (1999) Ternary rhythm and the *Lapse constraint. *Phonology* 16: 273–330. © 1999 by Cambridge University Press.

9 Kenstowicz, Michael (1996) Quality-sensitive stress. *Rivista di Linguistica* 9: 157–87.

10 Baković, Eric (1998) Unbounded stress and factorial typology. In *RuLing Papers 1: Working Papers from Rutgers University*, ed. Ron Artstein and Madeline Holler, pp. 15–28. New Brunswick, NJ: Department of Linguistics, Rutgers University. Reprinted by permission of Eric Baković.

11 Alderete, John (1999) Faithfulness to prosodic heads. In *The Derivational Residue in Phonological Optimality Theory*, ed. Ben Hermans and Marc van Oostendorp, pp. 29-50. Amsterdam: John Benjamins. Reprinted with kind permission of John Benjamins Publishing Company, Amsterdam/Philadelphia. www.benjamins.com.

12 Yip, Moira (1999) Feet, tonal reduction and speech rate at the word and phrase level in Chinese. In *Phrasal Phonology*, ed. René Kager and Wim Zonneveld, pp. 171–94. Nijmegen: Nijmegen University Press. Reprinted by permission of Nijmegen University Press.

13 Myers, Scott (1997) OCP effects in Optimality Theory. *Natural Language and Linguistic Theory* 15: 847–92. © 1997 by Kluwer Academic Publishers. Reprinted with kind permission of Kluwer Academic Publishers.

14 Pater, Joe (1999) Austronesian nasal substitution and other NC̥ effects. In *The Prosody–Morphology Interface*, ed. René Kager, Harry van der Hulst, and Wim Zonneveld, pp. 310–43. Cambridge: Cambridge University Press. © 1999 by Cambridge University Press.

15 Hayes, Bruce (1999) Phonetically driven phonology: the role of Optimality Theory and inductive grounding. In *Functionalism and Formalism in Linguistics*, vol. 1: *General Papers*, ed. Michael Darnell, Frederick J. Newmeyer, Michael Noonan, Edith Moravcsik, and Kathleen Wheatley, pp. 243–85. Amsterdam: John Benjamins. Reprinted with kind permission of John Benjamins Publishing Company, Amsterdam/Philadelphia. www.benjamins.com.

16 Beckman, Jill (1998) *Positional Faithfulness*. Doctoral dissertation. Amherst, MA: University of Massachusetts, Amherst. Published (1999) by Garland Publishing in Outstanding Dissertations in Linguistics Series. Reprinted by permission of Jill Beckman.

17 Lombardi, Linda (1999) Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17: 267–302. © 1999 by Kluwer Academic Publishers. Reprinted with kind permission of Kluwer Academic Publishers.

18 Zoll, Cheryl (1998) Positional asymmetries and licensing. Unpublished manuscript. Cambridge, MA: MIT. Reprinted by permission of Cheryl Zoll.

19 Padgett, Jaye (1995) Partial class behavior and nasal place assimilation. In *Proceedings of the 1995 Southwestern Workshop on Optimality Theory*

(SWOT), ed. Keiichiro Suzuki and Dirk Elzinga. Tucson, AZ: Department of Linguistics, University of Arizona. Reprinted by permission of Jaye Padgett.

20 Alderete, John (1997) Dissimilation as local conjunction. In *Proceedings of the North East Linguistic Society 27*, ed. Kiyomi Kusumoto, pp. 17–32. Amherst, MA: GLSA Publications. Reprinted by permission of John Alderete.

21 Kirchner, Robert (1996) Synchronic chain shifts in Optimality Theory. *Linguistic Inquiry* 27: 341–50. © 1996 by the Massachusetts Institute of Technology.

22 Benua, Laura (1997) *Transderivational Identity: Phonological Relations between Words*. Doctoral dissertation: University of Massachusetts, Amherst. Published (2000) by Garland Publishing in Outstanding Dissertations in Linguistics Series. Reprinted by permission of Laura Benua.

23 Rubach, Jerzy (2000) Backness switch in Russian. *Phonology* 17: 39–64. © 2000 by Cambridge University Press.

24 McCarthy, John J. and Prince, Alan (1993) Generalized alignment. In *Yearbook of Morphology*, ed. Geert Booij and Jaap van Marle, pp. 79–153. Dordrecht: Kluwer. © 1993 by Kluwer Academic Publishers. Reprinted with kind permission of Kluwer Academic Publishers.

25 Selkirk, Elisabeth (1996) The prosodic structure of function words. In *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, ed. James L. Morgan and Katherine Demuth, pp. 187–214. Mahwah, NJ: Lawrence Erlbaum Associates.

26 McCarthy, John J. and Prince, Alan (1994) The emergence of the unmarked: Optimality in prosodic morphology. In *Proceedings of the North East Linguistic Society 24*, ed. Mercè Gonzàlez, pp. 333–79. Amherst, MA: GLSA Publications. Reprinted by permission of John McCarthy and Alan Prince.

27 de Lacy, Paul (2002) Maximal words and the Māori passive. Reproduced by permission of Paul de Lacy.

28 Mascaró, Joan (1996) External allomorphy as emergence of the unmarked. In *Current Trends in Phonology: Models and Methods*, ed. Jacques Durand and Bernard Laks, pp. 473–83. Salford, Manchester: European Studies Research Institute, University of Salford. Reprinted by permission of Joan Mascaró.

29 Łubowicz, Anna (2002) Derived environment effects in Optimality Theory. *Lingua* 112: 243–80. © 2002, with permission from Elsevier.

30 Itô, Junko, Mester, Armin, and Padgett, Jaye (1995) Licensing and underspecification in Optimality Theory. *Linguistic Inquiry* 26: 571–614. © 1995 by the Massachusetts Institute of Technology.

31 Inkelas, Sharon, Orgun, Orhan, and Zoll, Cheryl (1997) The implications of lexical exceptions for the nature of grammar. In *Derivations and Constraints in Phonology*, ed. Iggy Roca, pp. 393–418. Oxford: Oxford University Press. Reprinted by permission of Sharon Inkelas, Orhan Orgun, and Cheryl Zoll.

32 Itô, Junko and Mester, Armin (1999) The phonological lexicon. In *The Handbook of Japanese Linguistics,* ed. Natsuko Tsujimura, pp. 62–100. Oxford: Blackwell. Reprinted by permission of Blackwell Publishing.

33 Anttila, Arto and Cho, Young-mee Yu (1998) Variation and change in Optimality Theory. *Lingua* 104: 31–56. © 1998, with permission from Elsevier.

Every effort has been made to trace copyright holders and to obtain their permission for the use of copyright material. The editor and publisher apologize for any errors or omissions in the above list and would be grateful if notified of any corrections that should be incorporated in future reprints or subsequent editions of this book.

Part I

The Basics

Chapter 1 | Alan Prince and Paul Smolensky

Optimality Theory: Constraint Interaction in Generative Grammar

Editor's Note

Optimality Theory first gained wide exposure from a course taught by Prince and Smolensky at the 1991 Summer Institute of the Linguistic Society of America. The earliest and still the most detailed exposition of the theory is their 1993 manuscript, an excerpt from which is here published for the first time. There has been much interest in this emerging theory; it has been the subject of a large and growing professional literature, an extensive electronic archive (http://roa.rutgers.edu), many courses and conference papers, and several textbooks. Although it was originally applied to phonology, the relevance of OT to topics in phonetics, morphology, syntax, sociolinguistics, psycholinguistics, and semantics has become increasingly apparent.

This chapter includes these excerpts: introductory material and motivation for the theory, including an analysis of Berber syllabification, drawn from sections 1 and 2 of Prince and Smolensky (P&S) (1993); an explanation of how constraints and constraint hierarchies evaluate candidates (section 5 of P&S 1993); the basic CV syllable theory with elaborations (section 6 and part of section 8 in P&S 1993); the theory of inventories and the lexicon (most of section 9 in P&S 1993). Readers may encounter sporadic references to other parts of P&S (1993): sections 3 and 4 on blocking and triggering (exemplified with Tongan stress, Tagalog infixation, Hindi stress, and Latin foot and word structure); section 7 on Lardil phonology; and section 10 on OT's relationships with functionalism, computation, Connectionism, Harmony Theory, and constraint-and-repair theories.

Readers approaching OT for the first time should begin with sections 1.2 and 2 of this chapter, followed by section 6, and then section 5. Readers can then go on to read the other parts of this chapter or other chapters in this book. Some natural pairings: the constraint HNUC in section 2 of this chapter re-emerges in stress theory in chapter 9; the CV syllable theory in section 6 of this chapter is studied from the perspectives of parsing and learning in chapters 4 and 5, respectively; the idea of faithfulness constraints (section 6.2.1) is generalized in chapter 3; emergence of

Excerpt (with minor revisions by the authors) from:
Prince, Alan and Smolensky, Paul (1993) *Optimality Theory: Constraint Interaction in Generative Grammar*. New Brunswick, NJ: Rutgers University Center for Cognitive Science. Technical report RuCCS-TR-2. [Available on Rutgers Optimality Archive, ROA-537.]

the unmarked is discussed briefly at the end of section 6.1 in this chapter and is the subject of chapter 26; lexicon optimization, which is discussed in section 9.3 of this chapter, is the topic of chapter 32.

[. . .]

1.2 Optimality

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

(1) $A \rightarrow B / C—D$

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.[1] The process can be schematically represented like this [the function H-eval, 'Harmonic Evaluation', determines the relative Harmony of the candidates]:

(2) Structure of Optimality-theoretic Grammar
  a. Gen (In_k) $\rightarrow$ $\{Out_1, Out_2, \ldots\}$
  b. H-eval (Out_i, $1 \leq i \leq \infty$) $\rightarrow$ Out_{real}

The grammar must define a pairing of underlying and surface forms, (input_i, 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 by scrutiny of output forms alone. A 'candidate' is an input–output pair, here formally encoded in what is called 'Out_i' in (2).

Gen contains information about the representational primitives and their universally irrevocable relations: for example, that the node σ may dominate a node *Onset* or a node μ (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 and others – 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 ⇆ 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 1988a, 1988b, 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-unto-α' 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 1991, 1993 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).[2]

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 paper, 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-α 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 find it. 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.
[. . .]

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 a syllable. One regularly encounters syllables of the shape *tK*, *rB*, *xZ*, *wL*, for example. (Capitalization represents nucleus-hood of consonants.) Table 1 provides illustrative examples, with periods used to mark syllable edges.[3]

Table 1

Nucleus type	Example	Morphology	Reference
voiceless stop	.ra.t**K**.ti.	ra-t-kti	1985: 113
voiced stop	.b**D**.dL. .ma.ra.t**G**t.	bddl ma=ra-t-g-t	1988: 1 1985: 113
voiceless fricative	.t**F**.tKt. .t**X**.zNt.	t-ftk-t t-xzn-t	1985: 113 1985: 106
voiced fricative	.tx**Z**.nakk^{w}.	t-xzn#nakk^{w}	1985: 113
nasal	.tz**M**t. .t**M**.zħ.	t-zmt t-mzħ	1985: 112 1985: 112
liquid	.t**R**.gLt.	t-rgl-t	1985: 106
high vowel	.**i**l.d**i**. .rat.l**u**lt.	i-ldi ra-t-lul-t	1985: 106 1985: 108
low vowel	.tR.b**a**.	t-rba	1985: 106

Dell and Elmedlaoui marshall 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.

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. tz**M**t — *t**Z**mt ‘*m* beats *z* as a nucleus’
  b. rat.l**u**lt — *ra.t**L**.w**L**.t ‘*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 *wL* 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 α.

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 1985), 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.[4] 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.[5]

The DEA is a rule, or rather a schema for rules, of exactly the classical type A → B / C—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:

(6) DEA in Action

Steps of the DEA	/ratlult/ 'you will be born'
Seek [X][+low,–cns] *& Build*	(**ra**)tlult
Seek [X][–low,–cns] *& Build*	(ra)t(**lu**)lt
Seek [X][+cns,+son,–nas]	*–blocked by FEC–*
Seek [X][+cns,+son,+nas]	—
Seek [X][–son,+cnt,+voi]	—
Seek [X][–son,+cnt,–voi]	—
Seek [X][–son,–cnt,+voi]	—
Seek [X][–son,–cnt,–voi] *& Build*	(ra)t(lu)(**lT**)[6]

(7) DEA in Action

Steps of the DEA	/txznt/ 'you sg. stored'
Seek [X][+low,–cns]	—
Seek [X][–low,–cns]	—
Seek [X][+cns,+son,–nas]	—
Seek [X][+cns,+son,+nas] *& Build*	tx(**zN**)t
Seek [X][–son,+cnt,+voi]	—
Seek [X][–son,+cnt,–voi] *& Build*	(**tX**)(zN)t
Seek [X][–son,–cnt,+voi]	—
Seek [X][–son,–cnt,–voi]	—

(8) DEA in Action

Steps of the DEA	/txznas/ ‘she stored for him’
Seek [X][+low,–cns] *& Build*	txz(**na**)s
Seek [X][–low,–cns]	—
Seek [X][+cns,+son,–nas]	—
Seek [X][+cns,+son,+nas]	*–blocked by FEC–*
Seek [X][–son,+cnt,+voi] *& Build*	t(**xZ**)(na)s
Seek [X][–son,+cnt,–voi]	—
Seek [X][–son,–cnt,+voi]	—
Seek [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?[7]

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)

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-α 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 α is *more* sonorous than segment β, 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 principles that define relative well-formedness 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.[8]

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.

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 (McCarthy & Prince 1993: §4). The nuclear constraint looks like this:[9]

(12) The Nuclear Harmony Constraint (Hnuc)
A higher sonority nucleus is more harmonic than one of lower sonority.
i.e. If $|x| > |y|$ then Nuc/x ≻ 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 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, 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 /ħaul-tn/ ‘make them (m.) plentiful’ (Dell & Elmedlaoui 1985: 110). Both Ons and Hnuc agree that the core syllable *ħa* should be formed: it has an onset as well as the best possible nucleus. Similarly, we must have a final syllable *tN*. 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 onsetted syllable with an inferior nucleus, as in *wL*. This situation can be perspicuously displayed in tabular form:[10]

(14) Constraint Inconsistency

Candidates /ħaul-tn/	Ons	Hnuc
~.wL.~		|l|
~.**u**l.~	*	|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, *.ħa.wL.tN*. 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 ≫ 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*:

(15) Constraint Tableau for Partial Comparison of Candidates from /ħaultn/

Candidates	Ons	Hnuc
☞ ~.wL.~		|l|
~.**u**l.~	* !	|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 .**u**l. 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 'overriding', 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(z**N**)t		n	optimal: onsetted, best available nucleus
txz(**N**)t	* !	n	no onset, Hnuc irrelevant
t(x**Z**)nt		z !	$\vert z\vert < \vert n\vert$
(t**X**)znt		x !	$\vert x\vert < \vert n\vert$
txz(n**T**)		t !	$\vert t\vert < \vert n\vert$

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.	* !	z	n	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.

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 §4 and §7 [omitted here – Ed.] 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. [omitted here – Ed.])

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.

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.

[. . .]

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, the other 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 Hɴᴜᴄ above. Similarly, tongue-height in vowels is neither harmonic nor disharmonic in isolation, but when the dimension of ATR (Advanced Tongue Root) 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

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

The goal of contextual markedness theory is to give content to the operator ⊗. Below in §8 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, 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	$\succ$	$\acute{a} \succ \acute{l}$	a nucleus filled by *a* is more harmonic than a nucleus filled by *l*
Domination hierarchy	Relative priority of well-formedness	$\gg$	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 domination 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 $\mathbb{C}$ generates marks of its own characteristic type, $*\mathbb{C}$. 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 σ 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 Ons(A) = (*Ons, *Ons, . . .) be a list containing one mark *Ons for each onsetless syllable in A. Thus for example Ons(*.txź.ńt.*) = (*Ons): the second, onsetless, syllable earns the parse *.txź.ńt.* its sole *Ons mark. (Here we use *ź* 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 Ons(A) and Ons(B) to see which contains fewer *Ons marks. We can notate this as follows:

$$\mathrm{A} \succ_{\textsc{Ons}}^{\text{parse}} \mathrm{B} \text{ iff } \textsc{Ons}(\mathrm{A}) \succ^{(*)} \textsc{Ons}(\mathrm{B})$$

where '$\succ_{\textsc{Ons}}^{\text{parse}}$' denotes comparison of entire parses and '$\textsc{Ons}(\mathrm{A}) \succ^{(*)} \textsc{Ons}(\mathrm{B})$' means 'the list Ons(A) contains fewer marks *Ons than the list Ons(B)'. (We will use the notation '(*)' as a mnemonic for 'list of marks'.) If the lists are the same length, then we write[11]

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

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.

Suppose we are given two lists of identical marks $*\mathbb{C}$; 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 $*\mathbb{C}$ 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 $*\mathbb{C}$ 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 * 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 ø, the null element.

Now let α and β be two lists of marks. We write $\alpha \succ^{(*)} \beta$ for 'α is more harmonic than β', which in the current context means 'α is *shorter* than β', 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 $*\mathbb{C}$, we adopt the following relation between single marks:

(96) Marks are Anti-harmonic

$$ø \succ^{*} *\mathbb{C}$$

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

$ø \approx^{*} ø$ and $*\mathbb{C} \approx^{*} *\mathbb{C}$

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

(97) Harmonic Ordering – Lists of Identical Marks

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

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

or

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

'$\beta \prec^{(*)} \alpha$' 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: α is shorter than β iff (if and only if) one of the following is true: (i) the first member of α is null and the first member of β is not (i.e., α is empty and β is not), or (ii) the list left over after removing the first member of α is shorter than the list left over after removing the first member of β.[12]

Now we can say precisely how Ons assesses the relative Harmony of two candidate parses, say *.tx́.zńt.* and *.txź.ńt.* 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:

.tx́.zńt. $\succ_{\text{Ons}}^{\text{parse}}$ *.txź.ńt.* because Ons(*.tx́.zńt.*) = ( ) $\succ^{(*)}$ (*Ons) = Ons(*.txź.ńt.*)

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

As another example:

.tx́.zńt. $\approx_{\text{Ons}}^{\text{parse}}$ *.txź.nt́.* because Ons(*.tx́.zńt.*) = ( ) $\approx^{(*)}$ ( ) = Ons(*.txź.nt́.*)

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

(99) Harmonic Ordering of Forms – Entire Parse, Single Constraint
Let $\mathbb{C}$ denote a constraint. Let A,B be two candidate parses, and let α,β be the lists of marks assigned them by $\mathbb{C}$:
$\quad$ $\alpha \equiv \mathbb{C}(A), \quad \beta \equiv \mathbb{C}(B)$
$\mathbb{C}$ by definition provides a Harmony order $\succ^{*}$ of the marks it generates. This order is extended to a Harmony order $\succ^{(*)}$ over lists of marks as follows:
$\quad$ $\alpha \succ^{(*)} \beta$ iff either:
$\qquad$ (i) $\mathrm{FM}(\alpha) \succ^{*} \mathrm{FM}(\beta)$
$\quad$ or
$\qquad$ (ii) $\mathrm{FM}(\alpha) \approx^{*} \mathrm{FM}(\beta)$ and $\mathrm{Rest}(\alpha) \succ^{(*)} \mathrm{Rest}(\beta)$
This order $\succ^{(*)}$ is in turn extended to a Harmony order over candidate parses (with respect to $\mathbb{C}$), $\succ_{\mathbb{C}}^{\text{parse}}$, as follows:
$\quad$ $A \succ_{\mathbb{C}}^{\text{parse}} B$ 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:

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

'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 *ń*. 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, HNUC(*.txź.ńt.*) = (*ń, ź*).

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 HNUC(A) and HNUC(B) are assumed sorted from most to least harmonic, this process is simply to compare the First Member of HNUC(A) with the First Member of 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 HNUC(A) and 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 ≡ *.tx́.zńt.* and B ≡ *.t́x.zńt.*

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

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

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

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

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

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

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

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

As it happens,

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

since both First Marks are *ń*, so we must discard the First Marks and examine the Rest, to determine whether

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

Here,

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