The Cambridge Handbook of Phonology

edited by Paul de Lacy

CAMBRIDGE www.cambridge.org/9780521848794

This page intentionally left blank

The Cambridge Handbook of Phonology

Phonology – the study of how the sounds of speech are represented in our minds – is one of the core areas of linguistic theory, and is central to the study of human language. This state-of-the-art handbook brings together the world's leading experts in phonology to present the most comprehensive and detailed overview of the field to date. Focusing on the most recent research and the most influential theories, the authors discuss each of the central issues in phonological theory, explore a variety of empirical phenomena, and show how phonology interacts with other aspects of language such as syntax, morphology, phonetics, and language acquisition. Providing a one-stop guide to every aspect of this important field, *The Cambridge Handbook of Phonology* will serve as an invaluable source of readings for advanced undergraduate and graduate students, an informative overview for linguists, and a useful starting point for anyone beginning phonological research.

PAUL DE LACY is Assistant Professor in the Department of Linguistics, Rutgers University. His publications include *Markedness: Reduction and Preservation in Phonology* (Cambridge University Press, 2006).

The Cambridge Handbook of Phonology

Edited by **Paul de Lacy**

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK
Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521848794

First published in print format 2007

ISBN-13 978-0-511-27605-7 eBook (Adobe Reader)
ISBN-10 0-511-27605-2 eBook (Adobe Reader)

ISBN-13 978-0-521-84879-4 hardback
ISBN-10 0-521-84879-2 hardback

Contents

Contributors

John D. Alderete, Assistant Professor, Department of Linguistics, Simon Fraser University.
Arto Anttila, Assistant Professor, Department of Linguistics, Stanford University.
Diana Archangeli, Professor, Department of Linguistics, University of Arizona.
Eric Baković, Assistant Professor, Linguistics Department, University of California, San Diego.
Ricardo Bermúdez-Otero, Lecturer, Department of Linguistics and English Language, University of Manchester.
Barbara Bernhardt, Associate Professor, School of Audiology and Speech Sciences, University of British Columbia.
Paula Fikkert, Associate Professor, Department of Dutch Language and Culture, Radboud Universiteit Nijmegen.
Stefan A. Frisch, Assistant Professor, Department of Communication Sciences and Disorders, University of South Florida.
Matthew Gordon, Associate Professor, Department of Linguistics, University of California, Santa Barbara.
Carlos Gussenhoven, Professor, Department of Linguistics, Radboud Universiteit Nijmegen and Queen Mary, University of London.
T. A. Hall, Assistant Professor, Department of Germanic Studies, Indiana University, Bloomington.
John Harris, Professor, Department of Phonetics and Linguistics, University College London.
René Kager, Professor, Utrecht Institute of Linguistics OTS (Onderzoeksinstituut voor Taal en Spraak), Utrecht University.
John Kingston, Professor, Department of Linguistics, University of Massachusetts Amherst.
Paul de Lacy, Assistant Professor, Department of Linguistics, Rutgers, The State University of New Jersey.

John J. McCarthy, Professor, Department of Linguistics, University of Massachusetts Amherst.

Alan Prince, Professor II, Department of Linguistics, Rutgers, The State University of New Jersey.

Douglas Pulleyblank, Professor, Department of Linguistics, University of British Columbia.

Keren Rice, Professor, Department of Linguistics, University of Toronto.

Joseph Paul Stemberger, Professor, Department of Linguistics, University of British Columbia.

Donca Steriade, Professor, Department of Linguistics and Philosophy, Massachusetts Institute of Technology.

Bruce Tesar, Associate Professor, Department of Linguistics, Rutgers, The State University of New Jersey.

Hubert Truckenbrodt, Assistent, Seminar für Sprachwissenschaft, Universität Tübingen.

Suzanne Urbanczyk, Associate Professor, Department of Linguistics, University of Victoria.

Adam Ussishkin, Assistant Professor, Department of Linguistics, University of Arizona.

Moira Yip, Professor, Department of Phonetics and Linguistics; Co-director, Centre for Human Communication, University College London.

Draga Zec, Professor, Department of Linguistics, Cornell University.

Acknowledgements

For a book of this size and scope it is probably unsurprising that many people contributed to its formation.

At Cambridge University Press, I owe Andrew Winnard a great deal of thanks. The idea for The Cambridge Handbook of Phonology was his, and it was a pleasure developing the project with him. My thanks also to Helen Barton for providing a great deal of editorial help throughout the process.

One of the most exhausting jobs was compiling, checking, and making consistent the seventeen hundred references. I am very grateful to Catherine Kitto and Michael O'Keefe for dealing with this task, and to Jessica Rett for contributing as well.

Of course, without the contributors, this volume would not exist. My thanks to them for meeting such difficult deadlines and responding so quickly to my queries.

A number of people commented on the initial proposal for this book, and every chapter was reviewed. My thanks go to: three anonymous reviewers, Crystal Akers, Akinbiyi Akinlabi, Daniel Altshuler, Eric Baković, Ricardo Bermúdez-Otero, Lee Bickmore, Andries Coetzee, José Elías-Ulloa, Colin Ewen, Randall Gess, Martine Grice, Bruce Hayes, Larry Hyman, Pat Keating, Martin Krämer, Seunghun Lee, John McCarthy, Laura McGarrity, Chloe Marshall, Nazarré Merchant, Jaye Padgett, Joe Pater, Alan Prince, Jessica Rett, Curt Rice, Sharon Rose, Elisabeth O. Selkirk, Nina Topintzi, Moira Yip, and Kie Zuraw. Of the reviewers, I must single out Kate Ketner and Michael O'Keefe: they carefully reviewed several of the articles each, provided the perspective of the book's intended audience, and also contributed a large number of insightful comments. There are also several times as many people again who 'unofficially' reviewed chapters for each author – my thanks to all those who in doing so contributed to this handbook.

Finally, I thank my colleagues and friends for advising and supporting me in this exhausting endeavour: Colin Ewen, Jane Grimshaw, John McCarthy, Alan Prince, Curt Rice, Ian Roberts, Moira Yip, and my colleagues

in the linguistics department at Rutgers. Finally, I thank my family – Mary and Reg for their unfailing support, and Sapphire and Socrates for their help with editing. Most of all I thank Catherine, whose encouragement and support were essential to my survival.

Introduction: aims and content

Paul de Lacy

Introduction

Phonological theory deals with the mental representation and computation of human speech sounds. This book contains introductory chapters on research in this field, focusing on current theories and recent developments.

1 Aims

This book has slightly different aims for different audiences. It aims to provide concise summaries of current research in a broad range of areas for researchers in phonology, linguistics, and allied fields such as psychology, computer science, anthropology, and related areas of cognitive science. For students of phonology, it aims to be a bridge between textbooks and research articles.

Perhaps this book's most general aim is to fill a gap. I write this introduction ten years after Goldsmith's (1995) *Handbook of Phonological Theory* was published. Since then, phonological theory has changed significantly. For example, while Chomsky & Halle's (1968) *The Sound Pattern of English* (*SPE*) and its successors were the dominant research paradigms over a decade ago, the majority of current research articles employ Optimality Theory, proposed by Prince & Smolensky (2004). Many chapters in this book assume or discuss OT approaches to phonology.

Another striking change has been the move away from the formalist conception of grammar to a functionalist one: there have been more and more appeals to articulatory effort, perceptual distinctness, and economy of parsing as modes of explanation in phonology. These are just two of the many developments discussed in this book.

2 Website

Supplementary materials for this book can be found on the website: http://handbookofphonology.rutgers.edu.

3 Audience and role

The chapters are written with upper-level undergraduate students and above in mind. As part of a phonology course, they will serve as supplementary or further readings to textbooks. All the chapters assume some knowledge of the basics of the most popular current theories of phonology. Many of the chapters use Optimality Theory (Prince & Smolensky 2004), so appropriate background reading would be, for example, Kager's (1999) textbook *Optimality Theory*, and for the more advanced McCarthy's (2002) *A Thematic Guide to Optimality Theory*.

Because it is not a textbook, reading the book from beginning to end will probably not prove worthwhile. Certainly, there is no single common theme that is developed step-by-step throughout the chapters, and there is no chapter that is a prerequisite for understanding any other (even though the chapters cross-reference each other extensively). So, the best use of this book for the reader is as a way to expand his/her knowledge of phonology in particular areas after the groundwork provided by a textbook or phonology course has been laid.

This book is also not a history of phonology or of any particular topics. While it is of course immensely valuable to understand the theoretical precursors to current phonological theories, the focus here is limited to issues in recent research.

4 Structure and content

The chapters in this book are grouped into five parts: (I) conceptual issues, (II) prosody, (III) segmental phenomena, (IV) internal interfaces, and (V) external interfaces.

The 'conceptual issues' part discusses theoretical concepts which have enduring importance in phonological theory: i.e. functionalist vs. formalist approaches to language, markedness theory, derivation, representation, and contrast.

Part II focuses on the segment and above: specifically prosodic structure, sonority, and tone. Part III focuses on subsegmental structure: features and feature operations. The chapter topics were chosen so as to cover a wide range of phenomena and fit in with the aims of phonology courses. However, while the areas in Parts II and III are traditionally considered distinct, the boundaries are at least fluid. For example, Gussenhoven

(Ch.11) observes that research on tone and intonation seems to be converging on the same theoretical devices, so the tone–intonation divide should not be considered a theoretically significant division. In contrast, some traditionally unified phenomena may consist of theoretically distinct areas: Archangeli & Pulleyblank (Ch.15) observe that there may be two separate types of harmony that require distinct theoretical mechanisms. Nevertheless, the division into discrete phenomena is inevitable in a book of this kind as in practice this is how they are often taught in courses and conceived of in research.

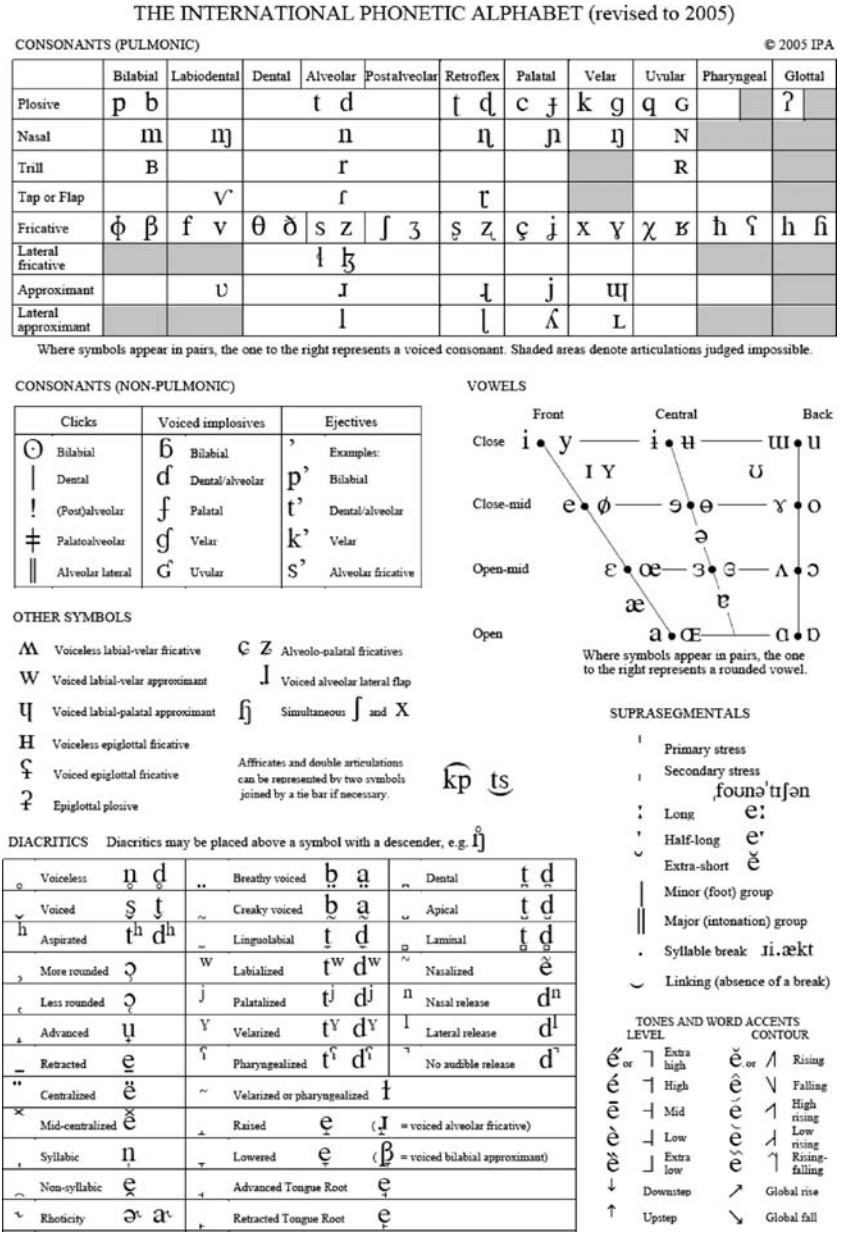
Part IV deals with 'internal interfaces' – the interaction of the phonological component with other commonly recognized modules – i.e. phonetics (Kingston Ch.17), syntax (Truckenbrodt Ch.18), and morphology (Ussishkin Ch.19 and Urbanczyk Ch.20).

Part V focuses on a variety of areas that do not fit easily into Parts I–IV. These include well-established areas such as diachronic phonology (Bermúdez-Otero Ch.21), areas that have recently grown significantly (e.g. language acquisition – Fikkert Ch.23) or have recently provided significant insight into phonological theory (e.g. free variation – Anttila Ch.22, learnability – Tesar Ch.24, phonological impairments – Bernhardt & Stemberger Ch.25).

Practical reasons forced difficult decisions about what to exclude. Nevertheless, as a number of phonologists kindly offered their views on what should be included I hope that the topics covered here manage to reflect the current concerns of the field.

While phonological research currently employs many different transcription systems, in this book an effort has been made to standardize transcriptions to the International Phonetic Alphabet (the IPA) wherever possible:
http://www2.arts.gla.ac.uk/IPA/index.html.

Chart of the International Phonetic Alphabet (revised 1993, updated 1996)



This chart is provided courtesy of the International Phonetics Association, Department of Theoretical and Applied Linguistics, School of English, Aristotle University of Thessaloniki, Thessaloniki 54124, GREECE.

1

Themes in phonology

Paul de Lacy

1.1 Introduction

This chapter has two aims. One is to provide a brief outline of the structure of this book; this is the focus of Section 1.1.1. The other – outlined in Section 1.1.2 – is to identify several of the major themes that run throughout.

1.1.1 Structure

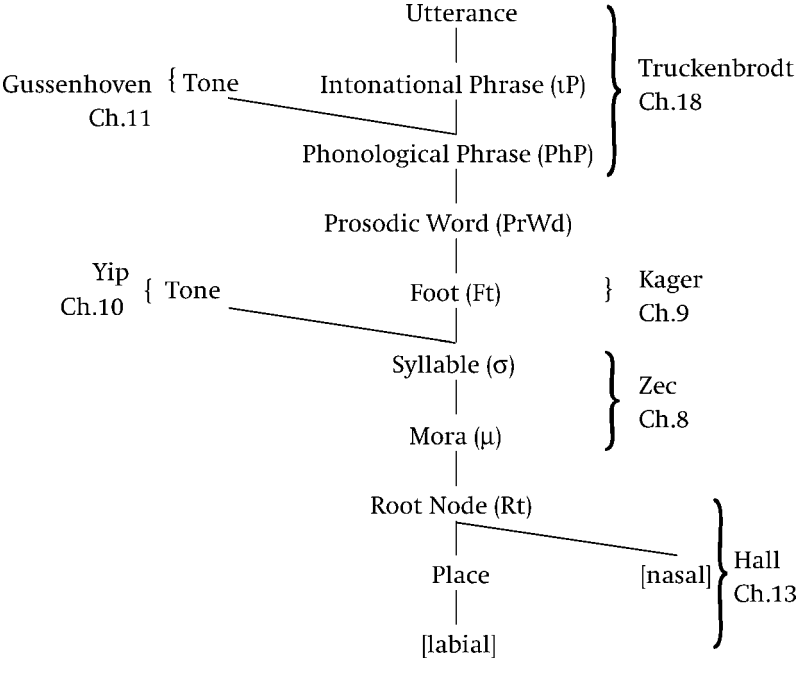
Several different factors have influenced the contents and structure of this Handbook. The topics addressed reflect theoretical concerns that have endured in phonology, but they were also chosen for pedagogical reasons (i.e. many advanced phonology courses cover many of the topics here). There were also 'traditional' reasons for some aspects of organization. While these concerns converge in the main, there are some points of disagreement. For example, there is a traditional distinction between the phonology of lexical tone and intonation, hence the separate chapters by Yip (Ch.10) and Gussenhoven (Ch.11). However, Gussenhoven (11.7) comments that theoretically such a division may be artificial.

Consequently, it is not possible to identify a single unifying theoretical theme that accounts for the structure of this book. Nevertheless, the topics were not chosen at random; they reflect many of the current concerns of the field. In a broad sense, these concerns can be considered in terms of representation, derivation, and the trade-off between the two. 'Representation' refers to the formal structure of the objects that the phonological component manipulates. 'Derivation' refers to the relations between those objects.

Concern with representation can be seen throughout the following chapters. Chomsky & Halle (1968) (*SPE*) conceived of phonological representation as a string of segments, which are unordered bundles of features. Since then, representation has become more elaborate. Below the segment, it is widely accepted that features are hierarchically organized (see discussion

and references in Hall Ch.13). Above the segment, several layers of constituents are now commonly recognized, called the 'prosodic hierarchy' (Selkirk 1984b). Figure (1) gives a portion of an output form's representation; it categorizes the chapters of this book in terms of their representational concerns. There is a great deal of controversy over almost every aspect of the representation given below – Figure (1) should be considered a rough expositional device here, not a theoretical assertion; the chapters cited should be consulted for details.

(1)



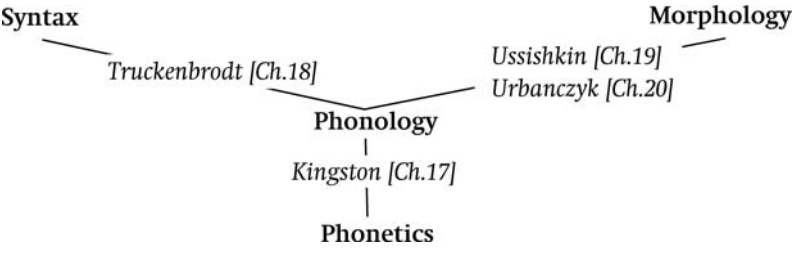
Harris (Ch.6) should be added to the chapters cited in (1); Harris' chapter is concerned with broader principles behind representation, including the notion of constituency, whether certain sub-constituents are phonologically prominent (i.e. headedness), and hierarchical relations.

Not represented in (1) is the interaction between constituents. For example, de Lacy (Ch.12) examines the interaction of tone, the foot, and segmental properties. Similarly, a part of Kager (Ch.9) is about the relation between the foot and its subconstituents. At the segmental level, three chapters are concerned with the interaction of segments and parts of segments: Baković (Ch.14), Archangeli & Pulleyblank (Ch.15), and Alderete & Frisch (Ch.16). For example, Baković's chapter discusses the pressure for segments to have identical values for some feature (particularly Place of Articulation).

Figure (2) identifies the chapters that are concerned with discussing the interaction of different representations. For example, Truckenbrodt (Ch.18) discusses the relation of syntactic phrases to phonological phrases. Ussishkin (Ch.19) and Urbanczyk (Ch.20) do the same for the relation of morphological

and phonological structure. Kingston (Ch.17) discusses the relation of phonological to phonetic structures.

(2)



There is also a 'derivational' theme that runs through the book chapters. McCarthy (Ch.5) focuses on evidence that there are relations between morphologically derived forms, and theories about the nature of those relations. Discussion of derivation has traditionally focused on the relation between input and output forms, and between members of morphological paradigms. However, the traditional conception of derivation has been challenged in Optimality Theory by McCarthy & Prince's (1995a, 1999) Correspondence Theory – the same relations that hold between separate derivational forms (i.e. input∼output, paradigmatic base∼derivative) also hold in the same output form between reduplicants and their bases; thus Urbanczyk's (Ch.20) discussion of reduplication can be seen as primarily about derivation, in this broadened sense.

Of course, no chapter is entirely about the representation of constituents; all discuss derivation of those constituents. In serialist terms, 'derivation of constituents' means the rules by which those constituents are constructed. In parallelist (e.g. Optimality Theoretic) terms, it in effect refers to the constraints and mechanisms that evaluate competing representations.

There is a set of chapters whose primary concerns relate to both representation and derivation: Prince (Ch.2), Gordon (Ch.3), Rice (Ch.4), and Steriade (Ch.7) discuss topics that are in effect meta-theories of representation and derivation. Gordon (Ch.3) examines functionalism – a name for a set of theories that directly relate to or derive phonological representations (and potentially derivations) from phonetic concerns. Rice (Ch.4) discusses markedness, which is effectively a theory of possible phonological representations and derivations. Steriade (Ch.7) discusses the idea of phonological contrast, and how it influences representation and derivation.

Rice's discussion of markedness makes the current tension between representation- and derivation-based explanations particularly clear. Broadly speaking, there have been two approaches to generalizations like "an epenthetic consonant is often [ʔ]". One assigns [ʔ] a representation that is different (often less elaborate) than other segments; the favouring of epenthetic [ʔ] over other segments is then argued to follow from general derivational principles of structural simplification. The other is to appeal to derivational principles such as (a) constraints that favour [ʔ] over every other segment and (b) no constraint that favours those other segments over

[ʔ]; [ʔ] need not be representationally simple (or otherwise remarkable) in this approach. These two approaches illustrate how the source of explanation – i.e. derivation and representation – is still disputed. The same issue is currently true of subsegmental structure – elaborated derivational mechanisms may allow simpler representational structures (Yip 2004).

Part V of this book contains a diverse array of phonological phenomena which do not fit easily into the themes of representational and derivational concerns. Instead, their unifying theme is that they are all areas which have been the focus of a great deal of recent attention and have provided significant insight into phonological issues; this point is made explicitly by Fikkert (Ch.23) for language acquisition, but also applies to the other areas: diachronic phonology (Bermúdez-Otero Ch.21), free variation (Anttila Ch.22), learnability (Tesar Ch.24), and phonological disorders (Bernhardt & Stemberger Ch.25). There are many points of interconnection between these chapters and the others, such as the evidence that phonological disorders and language acquisition provide for markedness.

Standing quite apart from all of these chapters is Prince (Ch.2). Prince's chapter discusses the methodology of theory exploration and evaluation.

In summary, no single theoretical issue accounts for the choice of topics and their organization in this book. However, many themes run throughout the chapters; the rest of this chapter identifies some of the more prominent ones.

1.1.2 Summary of themes

One of the clearest themes seen in this book is the influence of Optimality Theory (OT), proposed by Prince & Smolensky (2004).[1] The majority of chapters discuss OT, reflecting the fact that the majority of recent research publications employ this theory and a good portion of the remainder critique or otherwise discuss it.[2] However, one of the sub-themes found in the chapters is that there are many different conceptions and sub-theories of OT, although certain core principles are commonly maintained. For example, some theories employ just two levels (the input and output), while others employ more (e.g. Stratal OT – McCarthy 5.4). Some employ a strict and totally ordered constraint ranking, while others allow constraints to be unranked or overlap (see Anttila 22.3.3 and Tesar 24.4 for discussion). Theories of constraints differ significantly among authors, as do conceptions of representation (see esp. Harris Ch.6).

Another theme that links many of the chapters is the significance of representation and how it contributes to explanation. The late 1970s and 1980s moved towards limiting the form of phonological rules and elaborating the representation by devices such as autosegmental association, planar segregation, lack of specification, and feature privativity. In contrast, Harris (6.1) observes that the last decade has seen increased reliance on constraint form and interaction as sources of explanation. Constraint

interaction as an explanatory device appears in many of the chapters. Section 1.3 summarizes the main points.

Section 1.4 discusses the increasing influence of Functionalism in phonology, a theme that is examined in detail by Gordon (Ch.3). Reference to articulatory, perceptual, and parsing considerations as a source of phonological explanation is a major change from the Formalist orientation of *SPE* and its successors. This issue recurs in a number of chapters, some explicitly (e.g Harris 6.2.2, Steriade 7.5), and in others as an implicit basis for evaluating the adequacy of constraints.

Of course, the following chapters identify many other significant themes in current phonological theory; this chapter focuses solely on the ones given above because they recur in the majority of chapters and are presented as some of the field's central concerns.

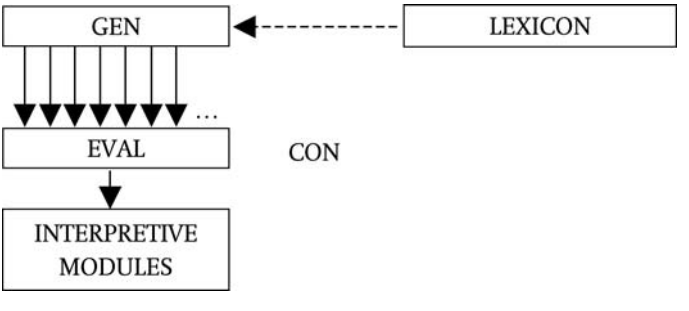
1.2 The influence of Optimality Theory

Optimality Theory is explicitly discussed or assumed in many chapters in this volume, just as it is in a great deal of current phonological research ('current' here refers to the time of writing – the middle of 2005). This section starts by reviewing OT's architecture and core properties. The following sections identify particular aspects that prove significant in the following chapters, such as the notion of faithfulness and its role in derivation in Section 1.2.1, some basic results of constraint interaction in Section 1.2.2, and its influence on conceptions of the lexicon in Section 1.2.3. The sections identify some of the challenges facing OT as well as its successes and areas which still excite controversy. The relation of OT to other theories is discussed in Section 1.2.4.

OT Architecture

OT is a model of grammar – i.e. both syntax and phonology (and morphology, if it is considered a separate component); the following discussion will focus exclusively on the phonological aspect and refer to the model in (3).

(3) *OT architecture*



For phonology, the GEN(erator) module takes its input either directly from the lexicon or from the output of a separate syntax module. GEN creates a possibly infinite set of candidate output forms; the ability to elaborate on the input without arbitrary restraint is called 'freedom of analysis'. In Prince & Smolensky's original formulation, every output candidate literally contained the input; to account for deletion, pieces of the input could remain unparsed (i.e. not incorporated into prosodic structure) which meant they would not be phonetically interpreted. Since McCarthy & Prince (1995a/1999), the dominant view is that output candidates do not contain the input, but are related to it by a formal relation called 'correspondence'; see Section 1.2.1 for details (cf. Goldrick 2000).

One significant restriction on GEN is that it cannot alter the morphological affiliation of segments ('consistency of exponence' – McCarthy & Prince 1993b). In practice it is common to also assume that GEN requires every output segment to be fully specified for subsegmental features, bans floating (or 'unparsed') features (except for tone – Yip 10.2.2, Gussenhoven 11.5.1), and imposes restrictions on the form of prosodic and subsegmental structure (though in some work they are considered violable – e.g. Selkirk 1995a, Crowhurst 1996, cf. Hyde 2002).

The EVAL(uator) module determines the 'winner' by referring to the constraints listed in CON (the universal constraint repository) and their language-specific ranking. Constraints are universal; the only variation across languages is (a) the constraints' ranking, and (b) the content of the lexicon. The winner is sent to the relevant interpretive component (the 'phonetic component' for phonology – Kingston Ch.17).

There are two general types of constraint: Markedness and Faithfulness. Markedness constraints evaluate the structure of the output form, while Faithfulness constraints evaluate its relationship to other forms (canonically, the input – see McCarthy Ch.5).[3] As an example, the Markedness constraint ONSET is violated once for every syllable in a candidate that lacks an onset (i.e. every syllable that does not start with a non-nuclear consonant – Zec 8.3.2). [ap.ki] violates ONSET once, while [a.i.o] violates it three times. The Faithfulness constraint I(nput)O(utput)-MAX is violated once for every input segment that does not have an output correspondent: e.g. /apki/ → [pi] violates IO-MAX twice (see Section 1.2.1 for details).

In each grammar the constraints were originally assumed to be totally ranked (although evidence for their exact ranking may not be obtainable in particular languages); for alternatives see Anttila (Ch.22). Constraints are violable; the winner may – and almost certainly will – violate constraints. However, the winner violates the constraints 'minimally' in the sense that for each losing candidate L, (a) there is some constraint K that favors the winner over L and (b) K outranks all constraints that favor L over the winner (a constraint 'favors' *x* over *y* if *x* incurs fewer violations of it than *y*); see Prince (2.1.1) for details.

Tableaux

The mapping from an underlying form to a surface form – a 'winner' – is represented in a 'tableau', as in (4). The aim here is to describe how to read a tableau, *not* how to determine a winner or establish a ranking: see Prince (2.1.1) for the latter.

The top left cell contains the input. The rest of the leftmost column contains candidate outputs. The winner is marked by the 'pointing hand'. C_3 outranks C_4 (shorthand: C_3 » C_4), as shown by the solid vertical line between them (C_1 outranks C_3, and C_2 outranks C_3, too). The dotted line between C_1 and C_2 indicates that no ranking can be *shown* to hold between them; it does not mean that there is no ranking.

Apart from the pointing hand, the winner can be identified by starting at the leftmost constraint in the tableau and eliminating a candidate if it incurs *more* violations than another contending candidate, where violations are marked by *s. For example, *cand*$_4$ incurs more violations than the others on C_1, so it is eliminated from the competition, shown by the '!'. C_2 likewise rules out *cand*$_3$. While *cand*$_4$ incurs fewer violations of C_3 than *cand*$_1$, it has already been eliminated, so its violations are irrelevant (shown by shading). C_3 makes no distinction between the remaining candidates as they both incur the same number of violations; it is fine for the winner to violate a constraint, as long as no other candidate violates the constraint *less*.

Another point comes out by inspecting this tableau: *cand*$_1$ incurs a proper subset of *cand*$_2$'s violation marks. Consequently, *cand*$_2$ can never win with any ranking of these constraints – *cand*$_1$ is a 'harmonic bound' for *cand*$_2$ (Samek-Lodovici & Prince 1999). Harmonic bounding follows from the fact that to avoid being a perpetual loser, a candidate has to incur fewer violations of some constraint for every other candidate; *cand*$_2$ doesn't incur fewer violations than *cand*$_1$ on any constraint.

(4) *A 'classic' tableau*

/input/	C_1	C_2	C_3	C_4
☞ (a) *cand*$_1$			*	*
(b) *cand*$_2$			*	* *!
(c) *cand*$_3$		*!		*
(d) *cand*$_4$	*!			

In some tableaux a candidate is marked with 💣 or ☜: these symbols indicate a winner that should not win – i.e. it is ungrammatical; in practical terms it means that the tableau has the wrong ranking or is considering the wrong set of constraints. In some tableaux, ☠ is used to mark a winner that is universally ungrammatical – i.e. it never shows up under any ranking; it indicates that there is a harmonic bound for the ☠-candidate.

The tableau form in (4) was introduced by Prince & Smolensky (2004) and is the most widely used way of representing candidate competition. Another method is proposed by Prince (2002a), called the 'comparative tableau'; it is used in this book by Prince (Ch.2), Baković (Ch.14), and Tesar (Ch.24).

The comparative tableau represents competition between pairs of candidates directly, rather than indirectly through violation marks. The leftmost column lists the winner followed by a competitor. A 'W' indicates that the constraint prefers the desired winner (i.e. the winner incurs fewer violations of that constraint than its competitor), a blank cell indicates that the constraint makes no preference, and an L indicates that the candidate favors the loser.

It is easy to see if a winner does in fact win: it must be possible to rearrange columns so that every row has at least one W before any L. Rankings are also easy to determine because on every row some W must precede all Ls. It's therefore clear from tableau (5) that both C_1 and C_2 must outrank C_3, and that C_1 must outrank C_4. It's also clear that it's not possible to determine the rankings between C_1 and C_2, C_2 and C_4, and C_3 and C_4 here. Harmonic bounding by the winner is also easy to spot: the winner is a harmonic bound for a candidate if there are only W's in its row (e.g. for the winner and $cand_2$ – it's harder to identify harmonic bounding between losers).

The comparative tableau format is not yet as widely used as the classic tableau despite having a number of presentational and – most importantly – analytical advantages over the classic type, as detailed by Prince (2002a).

(5) *A comparative tableau*

/input/	C_1	C_2	C_3	C_4
winner~$cand_2$				W
winner~$cand_3$		W	L	
winner~$cand_4$	W		L	L

Comparative tableaux can be annotated further if necessary: *e* can be used instead of a blank cell, and subscript numbers can indicate the number of violations of the loser in a particular cell (or even the winner's vs. loser's violations). The winner need not be repeated in every row: the top leftmost cell can contain the input→winner mapping, or the second row can contain the winner and its violations and the other rows can list the losers alone (i.e. just '~ *loser*' instead of '*winner*~*loser*').

Bernhardt & Stemberger (1998) propose another way of representing tableaux that is similar to the classic form; see Chapter 25 for details.

Core principles

Prince & Smolensky (2004) identify core OT principles for computing input→output mappings, including freedom of analysis, parallelism, constraint violability, and ranking. As they observe, many theories of CON and representation are compatible with these principles. Consequently, a

great deal of work in OT has focused on developing a theory of constraints; for proposals regarding other principles, see Section 1.2.4.

The dominant theories before OT – *SPE* and its successors – employed rules and a 'serial' derivation. For them, the input to the phonological component underwent a series of functions ('rules') that took the previous output and produced the input to the next until no more rules could apply. For example, /okap/ would undergo the rule $C \rightarrow \emptyset / \_]_\sigma$ to produce [oka] which would then serve as the input to the rule $V \rightarrow \emptyset /_\sigma [\_$ to produce [ka]. Rule-based derivation is described in detail in McCarthy (Ch.5). In contrast, the winner in OT is determined by referring to the constraint hierarchy and by comparison with (in principle) the entire candidate set (McCarthy & Prince 1993b:Ch.1§1).

Certainly, other theories had and have since proposed such concepts as constraints and two- or three-level grammars (e.g. Theory of Constraints and Repair Strategies – Paradis 1988; Harmonic Phonology – Goldsmith 1993a, Two-level Phonology – Koskenniemi 1983, Karttunen 1993; Declarative Phonology – Scobbie 1992, Coleman 1995, Scobbie, Coleman, and Bird 1996). However, OT's combination of these ideas and the key notions of constraint universality, ranking, and violability proved to have wide and almost immediate appeal.

The following sections discuss aspects of the theory that recur or are assumed in many of the following chapters. Section 1.2.1 discusses derivation, correspondence, and faithfulness. Section 1.2.2 discusses the form of the constraint component CON and some important constraint interactions while Section 1.2.3 examines OT's influence on the concept of the lexicon. Section 1.2.4 discusses the several different versions of OT that currently exist and their relation to other extant phonological theories.

1.2.1 Derivation and faithfulness

A concept that recurs throughout the following chapters is 'faithfulness' – it is discussed explicitly by McCarthy (Ch.5) and faithfulness constraints are used in many of the discussions of empirical phenomena.

In *SPE* and the theories that adopted its core aspects of rules and rule-ordering, there is no mechanism that requires preservation of input material. If input /abc/ surfaces as output [abc], the similarity is merely an epiphenomenon of rule non-application: either all rules fail to apply to /abc/, or the rules that apply do so in such a way as to inadvertently produce the same output as the input.

McCarthy & Prince (1995a, 1999) propose a reconceptualization of identity relations. Segments in different forms can stand in a relation of 'correspondence'. For example, the segments in an input $/k_1æ_2t_3/$ and winning faithful output $[k_1æ_2t_3]$ are in correspondence with one another, where subscript numerals mark these relations. Equally, the segments in an unfaithful pair, $/k_1æ_2t_3/ \rightarrow [d_1ɔ_3g_2]$, still correspond with one another,

even though in this case two segments have metathesized and all have undergone drastic featural change. In keeping with 'freedom of analysis', correspondence relations can vary freely among candidates. For example, input /$k_1æ_2t_3$/ has the outputs [$k_1æ_2t_3$], [$k_1æ_2$] (deletion of /t/), [$k_1æ_2t_3i$] (epenthesis of [i]), [$k_1t_3æ_2$] (metathesis of /æt/), [k_1æ$_{2,3}$] (coalescence of /æ/ and /t/ to form [æ]), and combinations such as [$t_3æ_2$] (metathesis of /æt/ and deletion of /k/) and forms that are harmonically bounded (i.e. can never win) such as [$k_3æ_3t_2$], and so on.

Constraints on faithfulness regulate the presence, featural identity, and linear order of segments. The ones proposed in McCarthy & Prince (1995a) that appear in this book are given in (6).

(6) *Faithfulness constraint summary (from McCarthy & Prince 1995a)*
  (a) Faithfulness constraints on segmental presence (e.g. Zec 8.3.2)
  MAX "Incur a violation for each input segment *x* such that *x* has no output correspondent." (Don't delete.)
  DEP "Incur a violation for each output segment *x* such that *x* has no input correspondent." (Don't epenthesize.)
  (b) Faithfulness constraints on featural identity (e.g. Steriade 7.4.3)
  IDENT[F] "Incur a violation for each input segment *x* such that *x* is [αF] and *x*'s ouput correspondent is [−αF]." (Don't change feature F's value.)
  (c) Faithfulness constraints on linear order (e.g. de Lacy 12.6)
  LINEARITY "For every pair of input segments *x,y* and their output correspondents *x',y'*, incur a violation if *x* precedes *y* and *y'* precedes *x'*." (No metathesis.)
  (d) Faithfulness constraints on one-to-many relationships (e.g. Yip 10.3.3)
  UNIFORMITY "Incur a violation for each output segment that corresponds to more than one input segment." (No coalescence.)

McCarthy & Prince (1995a, 1999) argue that correspondence relations can also hold within candidate outputs, specifically between reduplicative morphemes and their bases. Consequently, the candidate [$\underline{p_1a_2}p_1a_2t_3a_4$], where the underlined portion is the reduplicant, indicates that the reduplicant's [p] corresponds to the base's [p], and the reduplicant's [a] to the base's. This proposal draws a direct link between the identity effects seen in input→output mappings and those in base-reduplicant relations. Other elaborations of faithfulness are discussed in Section 1.2.2.

Parallelism

Faithfulness relates to the concept of parallelism: there is essentially a 'flat derivation' with the input related directly to output forms. As the chapters show, a lot of the success and controversy over parallelism arises in 'local' and 'non-local' interactions. One success is in its resolution of ordering paradoxes found in rule-based approaches. For example,

Ulithian's reduplication of /xas/ surfaces as [kakkasi] (Sohn & Bender 1973:45). Coda consonants assimilate to the following consonant, preventing the output from being *[xasxasi]. However, the form does not become the expected *[xaxxasi] because [xx] is banned. Instead, the resulting output is [kakkasi] – this form avoids [xx], satisfies the conditions on codas, and at the same time ensures that the reduplicant is as similar to the base as possible by altering the base's consonant from /x/ to [k].

The ordering paradox can be illustrated by a serialist rule-based analysis in (7). For the reduplicant to copy the base's [k] in [kakkasi], copying would have to be ordered *after* gemination and consequent fortition; however, reduplication *creates* the environment for gemination and fortition.

(7) *A serialist approach to Ulithian reduplication*

INPUT:	/RED+xasi/
(a) REDUPLICATION:	xas.xa.si
(b) GEMINATION:	xax.xa.si
(c) [xx] FORTITION:	*[xak.ka.si]

In contrast, Correspondence Theory (CT) provides an explanation by positing an identity relationship between the base and reduplicant. In tableau (8), CODACOND requires a coda consonant to agree with the features of the following consonant (after Itô 1986). *[xx] bans geminate fricatives. To force the input /x/ to become [k], both CODACOND and *[xx] must outrank IO-IDENT[continuant], a constraint that requires input-output specifications for continuancy to be preserved. Together, CODACOND and *[xx] favor the candidates with a [kk] – i.e. the winner [kak-kasi] and loser *[xak-kasi]. The crucial distinction between these two is that [kak-kasi]'s reduplicant copies its base's continuancy better than *[xak-kasi]'s. In short, the reason that [kak-kasi] wins is due to a direct requirement of identity between base and reduplicant (cf. discussion in Urbanczyk 20.2.6).

(8) *Ulithian reduplication in OT*

/RED-xasi/	CODA COND	*[xx]	BR-IDENT [continuant]	IO-IDENT [continuant]
(a) kak-kasi~xas-xasi	W		L	L
(b) kak-kasi~xax-xasi		W	L	L
(c) kak-kasi~xak-kasi			W	

Global conditions

Other aspects of faithfulness and parallelism have resulted in a great deal of controversy. One involves 'locality of interaction': a rule/constraint seems to apply at several places in the derivation (globality) or only once (opacity).

'Global rules' or 'global conditions' are discussed in detail by Anderson (1974): global conditions recur throughout a serial derivation. An example

that I am familiar with is found in Rarotongan epenthesis (Kitto & de Lacy 1999). There is a ban on [ɾi] sequences, and this ban recurs throughout the derivation. So, while the usual epenthetic vowel is [i] (e.g. [karaːti] 'carrot', [meneti] 'minute', [naeroni] 'nylon'), to avoid a [ɾi] sequence the epenthetic vowel after [ɾ] is a copy of the preceding one: e.g. [peːɾe] 'bail', [ʔamaɾa] 'hammer', [poːɾo] 'ball', [vuɾu] 'wool'. In serialist terms, the condition on [i] epenthesis seems straightforward: $\emptyset \rightarrow$ [i]/$C^{[-\text{rhotic}]}$__#, followed by a rule $\emptyset \rightarrow V_i$ /V_iɾ__#. The problem is that the ban on [ɾi] 'recurs' in the context [. . .iɾ]: if copying the vowel would result in a [ɾi] sequence, [a] is epenthesized as a last resort (e.g. [piɾa] 'bill', *[piɾi]). Consequently, the second rule needs to be reformulated as $\emptyset \rightarrow V_i$ /[¬i]ɾ__#, followed by $\emptyset \rightarrow$ [a] elsewhere. These rules miss the point entirely: there is a constraint on [ɾi] sequences that continually guides epenthesis throughout the derivation.

In OT, global conditions are expressed straightforwardly. A constraint on [ɾi] sequences outranks the constraints that would permit [ɾi]. The constraint M(¬i) is a shorthand for the constraints that favour [i] over all other vowels; AGREE(V) requires vowels to harmonize (Baković Ch.14, Archangeli & Pulleyblank Ch.15). In tableau (9), *[ɾi] is irrelevant because there is no [ɾ]; so the constraint M(¬i) favours [i] as the epenthetic vowel. In tableau (10), *[ɾi] blocks the epenthesis of [i], so the 'next best' option is taken – vowel harmony; this is one of the situations in which *[ɾi] blocks epenthesis. Tableau (11) illustrates the other: when harmony would produce an [ɾi] sequence, it is blocked and [a] is epenthesized instead.

(9) *Epenthesize [i] after non-[ɾ]*

/menet/	*ɾi	M(¬i)	AGREE(V)
(a) meneti~menete		W	L
(b) meneti~meneta		W	

(10) *. . . except when [i] epenthesis would result in [ɾi], then copy*

/vuːɾ/	*ɾi	M(¬i)	AGREE(V)
(a) vuːɾu ~vuːɾi	W	L	W
(b) vuːɾu ~vuːɾa			W

(11) *. . . unless copying would create [ɾi], in which case epenthesize [a]*

/piːɾ/	*ɾi	M(¬i)	AGREE(V)	M(¬a)
(a) piːɾa ~piːɾi	W	L	L	W
(b) piːɾa~piːɾe				W

Opacity

OT's success in dealing with global rules raises a problem. In a sense, the opposite of a global rule is one that applies in only one place in the derivation but not elsewhere, even when its structural description is met. Such cases are called 'opaque' and can be broadly characterized as cases where output conditions are not surface true. For example, an opaque version of Rarotongan epenthesis would have *ɾi apply only to block default [i]-epenthesis after [ɾ]; it would not block harmony, so allowing /piɾ/→[piɾi̲]. As McCarthy (5.4) discusses opacity in detail, little will be said about the details here (also see Bermúdez-Otero 21.3.2 for an example). Suffice to say that it is perhaps the major derivational issue that has faced OT over the past several years and continues to attract a great deal of attention. It has motivated a number of theories within OT, listed in McCarthy (Ch.5), and a number of critiques (e.g. Idsardi 1998, 2000). It is only fair to add that while opacity is seen as a significant challenge for OT, it also poses difficulties for a number of serialist theories: McCarthy (1999, 2003c) argues that serialist theories allow for unattested types of opaque derivation, where the input undergoes a number of rules that alter its form only for the output to end up identical to the input (i.e. 'Duke of York' derivations).

In summary, McCarthy & Prince's (1995a, 1999) theory that there is a direct requirement of identity between different derivational forms and even within forms has resulted in many theoretical developments and helped identify previously unrecognized phonological regularities. The opacity issue remains a challenge for OT, just as ordering paradoxes and global conditions pose problems for serialist rule-based frameworks.

1.2.2 Constraints and their interaction

Like many of the chapters in this book, a great deal of recent phonological research has been devoted to developing a theory of constraints. This Section discusses the basic constraint interactions and subtypes of faithfulness constraint that appear in the following chapters. The form of markedness constraints is intimately tied to issues of representation and Formalist/Functionalist outlook; these are discussed in Section 1.3 and Section 1.4 respectively.

Faithfulness

Many of the chapters employ faithfulness constraints that are elaborations of those in (6), both in terms of their dimension of application and environment-specificity.

McCarthy & Prince (1995a,1999) proposed that faithfulness relations held both on the input-output (IO) dimension and between bases and their reduplicants (BR) (see Urbanczyk Ch.20) for more on BR faithfulness). In its fundamentals, McCarthy & Prince's original conception of faithfulness

relations have remained unchanged: i.e. the core ideas of regulating segmental presence, order, and identity are still at the core of faithfulness. However, the dimensions over which faithfulness has been proposed to apply have increased. Correspondence relations within paradigms have been proposed by McCarthy, (1995, 2000c, 2005) and Benua (1997) (see McCarthy 5.5), from inputs to reduplicants by Spaelti (1997), Struijke (2000a/2002b) and others cited in Urbanczyk (20.2.6), and correspondence relations within morphemes have been explored by Kitto & de Lacy (1999), Hansson (2001b), and Rose & Walker (2004) (see Archangeli & Pulleyblank 15.3).

Others have proposed that there are environment-specific faithfulness constraints. For example, Beckman's (1997, 1998) 'positional faithfulness' theory proposes that constraints can preserve segments specifically in stressed syllables, root-initial syllables, onsets, and roots (also see Casali 1996, Lombardi 1999). For example, ONSET-IDENT[voice] is violated if an onset segment fails to preserve its underlying [voice] value, as in /aba/ → [a.pa] (but not /ab/→ [ap]) (see e.g. Steriade 7.4.3, Baković 14.4.3). There is currently controversy over whether positional faithfulness constraints are necessary, or whether their role can be taken over by environment-specific markedness constraints (Zoll 1998). For further elaborations on the form of faithfulness constraints in terms of environment, see Jun (1995, 2004), Steriade (2001b), and references cited therein.

In addition, some work seeks to eliminate particular faithfulness constraints, such as Keer (1999) for UNIFORMITY and Bernhardt & Stemberger (1998) and (25.3.4) for DEP.

A significant controversy relates to segment- and feature-based faithfulness. In McCarthy & Prince's (1995a) proposal, only segments could stand in correspondence with each other; a constraint like IDENT[F] then regulates featural identity as a property of a segment. In contrast, Lombardi (1999) and others have proposed that features can stand directly in correspondence – a constraint like MAX[F] requires that every input feature have a corresponding output feature. The difference is that the MAX[F] approach allows features to have a life of their own outside of their segmental sponsors. Consequently, the mapping /pa/→[a] does not violate IDENT[labial], but does violate MAX[labial]. For tone, MAX[Tone] constraints seem to be necessary (Yip 10.3, Myers 1997b), but for segments, it is common to use IDENT[Feature]. For critical discussion, see Keer (1999:Ch.2), Struijke (2000a/2002b:Ch.4), de Lacy (2002a§6.4.2), and Howe & Pulleyblank (2004).

Interactions of markedness and faithfulness

The source of much phonological explanation in OT derives from constraint interaction. At its most basic, the interaction of faithfulness and markedness determines whether input segments survive intact in the output (e.g. FAITH(α) » *α) or are eliminated (*α » FAITH(α)). In constraint terms, this is putting it fairly crudely: there are subtleties of constraint

interaction that can prevent elimination of underlying segments in different contexts. For example, Steriade (7.4.3) shows how the general ranking *βγ » *α » IDENT[α] prevents an otherwise general /α/→[β] mapping before γ (i.e. 'allophony').

One theme that the chapters here lack is explicit discussion of constraints on inputs. This is because interactions of faithfulness and markedness constraints preclude the need for restrictions on the input ('richness of the base' – Prince & Smolensky 2004: Sec. 9.3). For example, there is no need to require that inputs in English never contain a bilabial click /ʘ/; the general ranking *ʘ » FAITH[ʘ] will eliminate clicks in all output environments.

Turning to more subtle consequences of constraint interaction, a number of the following chapters employ a consequence of OT: the decoupling of rule antecedents and consequents. A rule like α→β describes both the 'problem' – i.e. α, and the 'solution' – i.e. β. In contrast, a constraint like *α identifies the problem without committing itself to any particular solution. *α could be satisfied by deleting α or altering α to β, for example. The proposal that the same constraint can have multiple solutions – both cross-linguistically and even in the same language – is called 'heterogeneity of process, homogeneity of target' (HoP-HoT – McCarthy 2002c: Sec 1.3.2).[4] Examples are found in various chapters: Baković (14.3) discusses the many ways that AGREE[F] can be satisfied, including assimilation, deletion, and epenthesis, with some languages employing more than one in different environments, Yip (10.3.3) shows how the OCP – a constraint on adjacent identical tones – can variously force tone deletion, movement, and coalescence in different languages, and de Lacy (12.6) shows how constraints that relate prosodic heads to sonority and tone can motivate metathesis, deletion, epenthesis, neutralization, and stress 'shift'.

While HoP-HoT has clearly desirable consequences in a number of cases, one current challenge is to account for situations where it over-predicts. For example, Lombardi (2001) argues that a ban on voiced coda obstruents can never result in deletion or epenthesis, only neutralization (e.g. such a ban can force /ab/ to become [ap] but never [a] or [a.bi̲]). This situation of 'too many solutions' is currently an area of increasing debate in OT (Lombardi 2001, Wilson 2000, 2001, Steriade 2001b, Pater 2003, de Lacy 2003b, Blumenfeld 2005).

Another consequence of constraint interaction is the Emergence of the Unmarked (TETU): a markedness constraint may make its presence felt in limited morphological or phonological environments (see e.g. Rice 4.5.1, 4.5.2, Urbanczyk 20.2.4). For example, a number of languages have only plain stops (e.g. Māori – Bauer 1993), so constraints against features like aspiration (*^{h}) must exist and in Māori outrank IDENT[^{h}]. In other languages where aspiration can appear fairly freely, IDENT[^{h}] outranks *^{h}. In contrast, in Cuzco Quechua *^{h} has an 'emergent' effect – while aspirated stops appear in roots, they do not appear at all in affixes. Beckman (1997§4.2.3)

shows that this pattern can be accounted for by the ranking Root-IDENT[h] » *h » IDENT[h], where Root-IDENT[h] is a positional faithfulness constraint that preserves aspiration in root segments only. Steriade (Ch.7) provides details.

TETU has provided insight into many areas of phonology. However, there are some challenging issues related to it. One is that in some languages, TETU results in a segment that is otherwise banned. For example, Dutch has an epenthetic [ʔ] in onsets, even though [ʔ] is otherwise banned in the language. For discussion, see Łubowicz (2003:Ch.5).

1.2.3 The lexicon

The chapters make both explicit and implicit assumptions about the form of the lexicon and the sort of information it provides in OT. The lexicon has been traditionally seen as the repository of 'unpredictable information' – it contains morphemes (or words) and their unpredictable properties, such as their morphological and syntactic categories, their phonological content, and their semantic content. Two ongoing issues with the lexicon are (a) where to store unpredictable information and (b) how much predictable information to store. In post-*SPE* phonology, the dominant view was to put all unpredictable information into the lexicon and to try to minimize predictable information. From the opposite point of view, Anderson (1992) proposed that at least some lexical items could effectively be expressed as rules.

Ussishkin (Ch.19) adopts a popular middle ground in OT, with some unpredictable aspects of morphemes implemented as constraints. For example, McCarthy & Prince (1993a) propose constraints such as ALIGN-L(*um*, stem), which requires the left edge of the morph of the Tagalog morpheme *um* to align with the left edge of a stem (i.e. be a prefix); this approach is discussed in detail by Ussishkin (19.3.2). So, whether a morpheme is prefixing or suffixing is not expressed in the lexicon as a diacritic that triggers a general concatenative rule (e.g. Sproat 1984), but as a morpheme-specific constraint.

The idea that unpredictable lexical information can be expressed as a rule/constraint is not due to OT, but OT has allowed expression of such information by constraints to be straightforward, and it is now widely assumed (cf. Horwood 2002). It is also debatable how much lexical information should be expressed as a constraint: Golston (1995) and Russell (1995) argue that even morphemes' phonological material should be introduced by constraint.

In *SPE*, as much predictable information was eliminated from the lexicon as possible and given by rule. For example, if medial nasal consonants always have the same place of articulation as the following consonant, pre-consonantal nasals in lexical entries were not specified for Place of Articulation. This idea was adapted in underspecification theories of the 1980s and 1990s. The explanatory power of *SPE* and its later rule-based

successors partly relied on the fact that the input to the phonology was restricted in predictable ways.

In contrast, Prince & Smolensky's (2004) principle of 'Richness of the Base' (RoB) forces this idea to be reconsidered. Because OT eschews constraints on input forms, a language's grammar must be able to account for every conceivable input, so whether underlying lexical forms lack predictable information or not becomes almost irrelevant with RoB. Consequently, a great deal of work in OT and in the chapters here assumes that lexical entries are fully specified for phonological information (cf. Itô et al. 1995, Inkelas et al. 1997, Artstein 1998). The irrelevance of the specification of predictable information in the lexicon does not indicate any greater level of complexity in OT. In fact, the principle it relies on – the lack of restrictions on inputs – has allowed resolution of some long-standing problems (e.g. the Duplication Problem – McCarthy 2002c§3.1.2.2).

Finally, a large amount of work in OT has re-evaluated the formal expression of morphological relatedness. As McCarthy (5.5) discusses, Correspondence Theory has been extended to account for phonological similarities among morphologically related words, such as the syllabic nasal in 'lighten' [laɪtn̩] and 'lightening' [laɪtn̩ɪŋ] (cf. 'lightning' [laɪtnəŋ], *[laɪʔn̩əŋ] – in the formal register of my dialect of New Zealand English) (e.g. Benua 1997). This issue is discussed more fully by McCarthy (5.5).

In short, the lexicon in OT is different in significant ways from the lexicon in previous work. Some unpredictable information has been moved out of the lexicon and expressed as constraints, and some predictable information is commonly assumed to remain in the lexicon. The formal expression of 'morphological relatedness' and paradigms has changed fundamentally as part of the development of Correspondence Theory; it is no longer necessary to appeal to a serial derivation to account for phonological similarities between morphologically related words.

1.2.4 OT theories and other theories

One point that emerges from surveying the chapters in this volume is that it is misleading to imply that there is a single unified theory of OT that everyone adheres to. It is more accurate to say that there is an OT framework and many OT sub-theories.

Almost every aspect of OT has been questioned. For example, McCarthy & Prince's (1995a, 1999) theory of GEN with Correspondence is fundamentally different from the Containment model of Prince & Smolensky (2004). There are also fundamentally different approaches to the constraint component CON: some view constraints from a Functionalist perspective and others from a Formalist one (see Section 1.3). In addition, some approaches see each constraint as independently motivated, while others attempt to identify general schemas that define large classes of constraints (e.g. McCarthy & Prince's 1993a ALIGN schema (Ussishkin 19.2.1), Beckman's

1997 positional faithfulness schema, and markedness schemas in a variety of other work). The concept of a totally ordered and invariant ranking has been questioned from several perspectives (see Anttila Ch.22 for details). Wilson (2000) proposes an EVAL that is fundamentally different from Prince & Smolensky's (2004) (cf. McCarthy 5.4, 2002b). McCarthy (2000b) examines – but does not advocate – a Serialist OT theory (also Rubach 2000). Finally, a number of proposals involving more than two levels have been put forward recently (see McCarthy 5.4).

In addition, the core principles of OT are compatible with aspects of other theories. For example, Harris & Gussmann (1998) combine representational elements of Government Phonology with OT. Some key features of the rule-based Lexical Phonology have been recast in an OT framework (see McCarthy 5.5).

In summary, there are many subtheories of OT, there are mixtures of OT and other theories' devices, and there also are a number of other theories that are the focus of current research (e.g. Government Phonology in Scheer 1998, 2004; Declarative Phonology – Coleman 1998, Bye 2003, and many others). Nevertheless, it is clear from the chapters here that Prince & Smolensky's (2004) framework has had a profound impact on the field and helped to understand and reconceptualize a wide variety of phonological phenomena.

1.3 Representation and explanation

Harris (6.1) observes that "recent advances in derivational theory have prompted a rethink of . . . representational developments." Comparison of the chapters in Goldsmith's (1995a) *Handbook* with the ones here underscores this point: here there is less appeal to specific representational devices and more reliance on constraints and their interaction to provide sources of explanation.

To give some background, in Autosegmental Phonology (Goldsmith 1976b, 1990) and Metrical Phonology (see Hayes 1995, Kager Ch.9 for references) the aim throughout the 1980s and early 1990s was to place as much of the explanatory burden as possible on representation with very few operations (e.g. relinking and delinking of association lines, clash and lapse avoidance). In contrast, constraint interaction in OT allows ways to analyze phonological phenomena that do not rely on representational devices.

Markedness and representation

An example is found in the concept of Markedness, which has been a central issue in phonological theory since the Prague School's work in the 1930s (Trubetzkoy 1939, Jakobson 1941/1968). It is the focus of Rice's chapter (Ch.4) in this handbook, and markedness theory is explicitly discussed in many others (e.g. Zec 8.5, de Lacy Ch.12, Fikkert Ch.23, Bernhardt & Stemberger 25.2.1).

'Markedness' refers to asymmetries in linguistic phenomena. For example, it has often been claimed that epenthesis can produce coronals, but never labials or dorsals (e.g. Paradis & Prunet 1991b and references cited therein). Coronals are therefore less marked than labials and dorsals, and this markedness status recurs in many other processes (e.g. neutralization). In general, phonological phenomena such as neutralization and epenthesis are taken to produce exclusively unmarked feature values.[5]

SPE's approach to markedness was to define feature values – *u* for unmarked and *m* for marked – which were interpreted by special 'marking conventions' which essentially filled in a phonetically interpretable value of '+' or '−' (*SPE*:Ch.9). *SPE*'s approach was therefore essentially representational: markedness follows from the form of feature values. After *SPE*, a more elaborate theory of representation and markedness developed in the Autosegmental Theory of representation (Goldsmith 1976a, 1990), and in theories of underspecification (e.g. Kiparsky 1982b, Archangeli 1984) and privativity (e.g. Lombardi 1991) (see Harris 6.3, Hall 13.2). The unmarked feature value was indicated by a lack of that feature; for example, coronals had no Place features at all (articles in Paradis & Prunet 1991b, Avery & Rice 1989, Rice 1996, also see Hall Ch.13). Coupled with the view that neutralization is feature deletion, the fact that neutralization produces unmarked elements is derived.

While the representational approach to markedness has continued in OT work (for recent work – Causley 1999, Morén 2003), Prince & Smolensky (2004) and Smolensky (1993) opened up an entirely different way of conceiving of the concept (its most direct precursor is in Natural Generative Phonology – Stampe 1973). Instead of relying on representation, constraint ranking and form is central: coronals are not marked because they are representationally deficient, but because all constraints that favour dorsals and labials over coronals are universally lower-ranked than those constraints that favor coronals over other segments: i.e. ‖ *DORSAL »» *LABIAL »» *CORONAL ‖, where '»»' indicates a ranking that is invariant from language to language. There is no need to appeal to the idea that coronals lack Place features in this approach: they are the output of neutralization because other options – labials and dorsals – are ruled out by other constraints (for examples of fixed ranking, see Zec 8.5, Yip 10.3.2, de Lacy 12.2.2).

The idea of universally fixed rankings is found in the opening pages of Prince & Smolensky (2004); its success at dealing with markedness hierarchies in the now famous case of Imdlawn Tashlhiyt Berber syllabification is probably part of the reason that OT's influence spread so quickly (see Zec 8.5.1 for discussion). Recent approaches to markedness in OT have rejected universally fixed rankings; they instead place restrictions on constraint form to establish markedness relations (see de Lacy Ch.12). However, the principle is the same: markedness relations are established by ranking and constraint form, not by representational devices.

The OT ranking/constraint form approach to markedness has been widely accepted in current work, but the representational theory also

remains popular: the two approaches are often even employed together. As discussed in Harris (Ch.6), the debate continues as to where the balance lies.

Representation in current theory

The chapters identify and exemplify a number of reasons why there was a shift towards explanation through constraint interaction. One function of representation was to express markedness; as explained above, from the first, Prince & Smolensky (2004) showed how to capture markedness effects with constraint interaction. Similarly, much of the theory of representation relied on, or at least employed, serial derivations. For example, assimilation was seen as a three-step process of delinking a feature, adding an association to a nearby feature, then deleting the stray feature (also see Harris 6.3.3). With a two-level approach to grammar, the concepts of delinking and reassociation have no clear counterpart (though see Yip Ch.10 and the discussion below).

In many of the chapters here, Correspondence Theory is used instead of representational devices. For example, reduplication was seen in Marantz (1982) and McCarthy & Prince (1986) as a series of associations followed by delinking due to a ban on crossed association lines; Urbanczyk (Ch.20) shows how reduplication can be analyzed using correspondence – another type of relation entirely. Representation was also relied on to express dependency relations. For example, if a feature F always assimilates whenever feature G does, then F was assumed to be representationally dependent on G. Harris (6.3.3) observes that Padgett's (2002) work shows that at least some dependency relations between features and classhood can be expressed through constraint interaction and do not rely on an explicit representational hierarchy of features (also see Yip 2004).

Of course, it is crucial for any theory of phonology to have a well-defined restrictive theory of representation. However, OT has allowed the burden of explanation to move from being almost exclusively representation-based to being substantially constraint-based.

In fact, while most recent work in OT has focused on constraint interaction, a good deal has examined or employed representational devices as a crucial part of explanation. For example, Beckman (1997, 1998) employs an OT version of Autosegmental phonology to deal with assimilation. Cole & Kisseberth (1994) propose Optimal Domains theory, which certainly relies less on representational devices than its predecessors but crucially refers to a representational notion of featural alignment. McCarthy (2004a) proposes a theory of representation that builds on autosegmental concepts. Interestingly, the representation of tone has been least affected by the move to OT. Very little has changed in representational terms: pre-OT notions such as multiply-linked (i.e. spread and contour) tones, floating tones, and tonal non-specification are commonly used in OT work – see Yip (Ch.10) for details.

One reason for the lack of in-depth discussion of representation is that it has become common to focus on constraint interaction and violations in OT work, while there has been less necessity to provide explicit definitions

of constraint form. An example is the AGREE[F] constraint (Lombardi 1999, Baković 2000), defined by Baković (14.1) as "Adjacent output segments have the same value of the feature *x*." The constraint is defined in this way because the definition aims to express the *effect* of the constraint (i.e. how it assigns violations) rather than providing a formal structural description. If one wishes to completely formalize the definition, though, it is necessary to deal with representational issues: what does the term "have the same value" mean? In formal terms, is this phrase necessarily expressed as a multiply-associated feature? Or can it be expressed through correspondence relations? These issues are receiving more attention in recent work.

In summary, much of the burden of explanation has shifted from representational devices to constraint interaction. However, many of the representational devices that were developed in the 1980s remain integral to current phonological analyses, as exemplified by the detailed prosodic structures used by Zec (Ch.8), Kager (Ch.9), Yip (Ch.10), Gussenhoven (Ch.11), and Truckenbrodt (Ch.18), and the feature structure discussed by Hall (13.2). As the authors discuss, justification for the structures remains despite the effects of constraint interaction.

1.4 Functionalism

Gordon (Ch.3) observes that "the last decade has witnessed renewed vigor in attempting to integrate functional, especially phonetic, explanations into formal analyses of phonological phenomena." Functionalist principles are discussed in many of the chapters in this book (including Rice 4.7, Harris 6.2.2, Zec 8.6, Steriade 7.3, Yip 10.4.2, Hall 13.2, Baković 14.4.1, Alderete & Frisch 16.3, Kingston 17.3, Bermúdez-Otero 21.4, Anttila 22.3.3, 22.4, Fikkert 23.2). This section provides some background to both Functionalist and Formalist approaches to phonology (also see McCarthy 2002c§4.4).

Gordon (Ch.3) identifies a number of core principles in Functionalist approaches to phonology. A central concept is expressed by Ohala (1972:289): "Universal sound patterns must arise due to the universal constraints or tendencies of the human physiological mechanisms involved in speech production and perception". Many researchers have advocated a Functionalist approach (e.g. Stampe 1973, Ohala 1972 et seq., Liljencrants & Lindblom 1972, Archangeli & Pulleyblank 1994, Bybee 2001 and many others), but it is only recently that Functionalist theories employing OT-like frameworks have gained a great deal of popularity, as documented by Gordon (Ch.3) (also see the articles in Gussenhoven & Kager 2001, Hume & Johnson 2001, and Hayes et al. 2004; Flemming 1995, Jun 1995, Boersma 1998, Kirchner 1998, 2001, Gordon 1999, 2002b, and many others). Research has focused on issues such as how concepts such as markedness are grounded in concepts of articulatory ease and perceptual distinctiveness, and how to express these influences in constraint form.

The property common to all current Functionalist approaches is the idea that phonological effects (especially markedness) are not due to innate constraints or constraint schemas. Instead, one Functionalist view (called 'Direct Functionalism' here) holds that constraints are constructed by mechanisms that measure articulatory effort and perceptual distinctiveness (and perhaps also parsing difficulty). Constraints are defined in units that directly record this effort and distinctiveness; consequently, the approaches use finely differentiated units (e.g. real numbers) not used in traditional conceptions of phonology (see Harris 6.2.2 for discussion, also Anttila 22.3.3 and Tesar 24.4).

Another view combines direct functionalism with the idea that the phonological component is limited in terms of its expressive power. In this view, constraints are constructed with reference to articulation and perception, but they must be expressed in terms of a small set of phonological primitives: i.e. "phonological constraints tend to ban phonetic difficulty in simple, formally symmetrical ways" (Hayes 1999§6.2). The phonological primitives may not be well-adapted to expressing phonetic categories, so there may be various mismatches.

Distinct from these views is the 'diachronic functionalist' approach (Ohala 1971 et seq., Blevins 2004). Blevins' approach in particular potentially allows the phonological component to generate virtually any sound pattern (Gordon 3.5). However, not every sound pattern survives diachronic transmission equally well. Consequently, markedness effects are due to the process of language learning, and explanation for diachronic change and synchronic processes are the same. Diachronic functionalism is discussed by Gordon (3.5), so will not be examined further here.

1.4.1 The Formalist approach

A great deal of current phonological work has its roots in Formalist approaches (see Chomsky 1966 for phonology specifically, Chomsky 1965 et seq., and more recently Hale & Reiss 2000b). In OT, the Formalist approach is responsible for the assumption that all constraints or constraint schemas are innate.

The Formalist approach does not necessarily rule out functional grounding in constraints. As Chomsky & Lasnik (1977§1.2) discuss, Formalist approaches can assume a 'species-level' functionalism: this is the idea that a particular constraint has been favoured in evolution because it helps with articulation, perception, or parsing. For example, Chomsky & Lasnik suggest that the syntactic constraint $^{*}[_{\text{NP}}$NP TENSE VP] is innate, and has survived because it simplifies parsing (p.436).

The implication of the Formalist approach for phonology is that derivation, representation, and constraints can have 'arbitrary' aspects – i.e. they may not directly aid (and could even act against) reduction of articulatory effort and increase in perceptual distinctiveness. However, it is not surprising to find that some (or even many) mechanisms or constraints do serve to

aid in articulation, perception, and processing; this functional grounding would be seen as following from 'species-level' adaptations or 'accident', through fortuitous random mutation or exaptation.

With 'species-level functionalism', it may seem that the Formalist and Functionalist approaches would have very similar effects. However, the difference resides in the Formalist possibility for arbitrary phonological structures, hierarchies, and constraints. For example, Zec (8.5) and de Lacy (12.2) employ the sonority hierarchy as a central part of their analyses of prosodic structure, yet determining the phonetic basis of sonority – and therefore its articulatory and perceptual value – has proven notoriously difficult (Parker 2002 and references cited therein). It seems that the sonority hierarchy is at least partially arbitrary (i.e. without functional motivation), and only partially adapted to aiding articulation and perception; this sort of mismatch is expected in the Formalist approach. Of course, the difficulty in identifying arbitrariness is that we may simply not be looking at the right articulatory, perceptual, or parsing property.

Also expected in the Formalist view is the idea that there could be arbitrary (and even functionally non-sensical) restrictions on phonological processes. An example is found in tone- and sonority-driven stress, discussed in de Lacy (Ch.12). Longer segments (e.g. long vowels, diphthongs) often attract stress, and there are plausible functional reasons for such attraction (Ahn 2000). In fact, this attraction may (partially) account for the attraction of stress to high sonority vowels like [a] because they typically have a longer inherent duration than low sonority vowels like [i], [u], and [ə]. However, in many languages there is a correlation between tone level and vowel duration: the lower the tone, the longer the vowel (e.g. Thai – Abramson 1962). Thus, low-toned [à] is longer than high-toned [á], and so on. If low tone increases duration, and stress is attracted to longer elements, functional reasoning should lead us to believe that stress will be attracted to low tone over high tone. However, this is *never* the case: stress always prefers high-toned vowels to low-toned ones. Of course, there may be some other functional reason for favouring high-toned stressed syllables, but given the fact that languages can vary as to which functional factor they favour (i.e. through ranking), it is surprising that *no* language favours stressed low-toned vowels over high-toned ones (cf. functional approaches to vowel inventories, where articulatory and perceptual factors can conflict, but one can take precedence over the other in particular languages).

To summarize, support for the Formalist view (with 'species-level functionalism') can be sought in phonological arbitrariness and Competence–Performance mismatches.

1.4.2 Challenges

Gordon (3.1) observes that one reason for the increase in Functionalist popularity is OT's formalism: OT can be easily adapted to expressing

gradient phenomena; it also provides a framework for expressing the concept of 'tendency' through constraint ranking. However, it is important to emphasize that OT is not an inherently Functionalist theory, and some Functionalist versions of OT depart significantly from Prince & Smolensky's (2004) proposals (e.g. versions of Stochastic OT – see discussion and references in Anttila 22.3.3, McCarthy 2002c: Sec. 4.4).

Another reason may be that the Formalist explanation for sound patterns is seen by some as insufficiently profound. For example, the fact that dorsals are more marked than coronals receives the explanation that *DORSAL universally outranks *CORONAL in a Formalist approach, and this universal ranking is innate. In other words, the constraint ranking is an axiom of the theory. Yet there is clearly a good articulatory reason for this ranking – dorsals require more articulatory effort than coronals (if effort is measured from rest position), and there may be perceptual reasons as well. A Functionalist approach makes a direct connection between the substantive facts and the formalism.

A further reason is skepticism about the ability of species-level functionalism to account for phonological facts. For example, how could the fixed ranking *DORSAL »» *CORONAL evolve? A fixed ranking *DORSAL »» *CORONAL would have to appear through a random mutation (or exaptation), then provide some advantage that a speaker who had to learn their ranking did not have (e.g. faster learning). Identifying the exact advantage (whether survival or sexual) is challenging. There may also be the issue of plausibility, though as Pinker & Bloom (1990) have observed, tiny advantages can have significant influence over time. On the other hand, natural selection is not the only force in biological evolution.

The problem that Formalist approaches face is not that they lack explanation, but that it is difficult to provide proof. Little is understood about the biology of phonological evolution, and so evolutionary arguments are hard to make (though see Hauser, Chomsky, and Fitch 2002 for discussion and references). Given the burgeoning popularity of Functionalist approaches, the onus currently seems to be on the Formalist approach to close the 'plausibility gap' and identify clearly testable predictions that differ from Functionalist ones.

There are also challenges for the Functionalist perspective. For the diachronic Functionalist view, one challenge is to account for cases where a diachronic change has no synchronic counterpart, and why there are unattested synchronic grammars which could easily be created by a series of natural diachronic changes (Kiparsky 2004). Mismatches also pose a challenge for the 'direct' Functionalist point of view, as do cases of arbitrariness (as in the sonority hierarchy), as all constraints and markedness hierarchies should be tied directly into Performance considerations.

Functionalist approaches have already had a significant impact on phonological theory. There are many works that explicitly advocate Functionalist principles (cited in Section 1.3 above). It is also commonplace in recent

publications to see a constraint's validity evaluated by whether it is related to a decrease in articulatory effort or helps in perception or parsing. For example, in a widely-used textbook, Kager (1999a:11) comments that "phonological markedness constraints should be *phonetically grounded* in some property of articulation and perception". Of course, a Formalist perspective does not accept the validity of such statements. In the immediate future, I think it is likely that the Functionalist perspective will continue to gain ground, but also that there will be increasing dialogue between the various Formalist and Functionalist approaches and increased understanding of the implications of Formalist tenets in phonology.

1.5 Conclusions

The preceding sections have attempted to identify some of the major theoretical themes that appear throughout the following chapters. Of course, there are many others in the following chapters that are not covered here (e.g. the role of contrast in phonology – Steriade Ch.7, Rice 4.6). Fikkert (23.1) comments that for language acquisition there has been an increase in research and resources, and Tesar (Ch.24) discusses the growing field of learnability. As detailed in the chapters in Part V, areas of phonology that have traditionally been under-studied or seen as peripheral (e.g. free variation – Anttila Ch.22) are having a significant influence on central issues in the field.

The chapters in this Handbook show that phonological theory has undergone enormous theoretical changes compared with ten years ago, and it continues to change rapidly. It is probably for this reason that none of the chapters in this book attempt to make predictions about the broad issues that will dominate phonology in the next ten years. Perhaps the only safe bet is that any prediction about the future of phonology will be wildly inaccurate.

Notes

My thanks to all those who commented on this chapter in its various incarnations: José Elías-Ulloa, Kate Ketner, John McCarthy, Nazarré Merchant, Michael O'Keefe, and Alan Prince.

1 Prince & Smolensky's manuscript was originally circulated in 1993. A version is available online for free at the Rutgers Optimality Archive (ROA): http://roa.rutgers.edu/, number 537.

2 From inspecting several major journals from 1998 to 2004, around three-quarters of the articles assumed an OT framework, and many of the others compared their theories with an OT approach.

3 There is currently an ambiguity in the term 'markedness'. In OT, 'markedness' refers to a type of constraint. 'Markedness' also refers to a concept of implicational or asymmetric relations between phonological segments and structures (see Section 1.3 and Rice Ch.4).
4 The opposite is identified and exemplified by Ketner (2003) as 'heterogeneity of target, homogeneity of process', where the same process is used to satisfy a number of different conditions.
5 There is a great deal of controversy over the role of markedness in phonology. For example, Blevins (2004) proposes that markedness effects can be ascribed to diachronic change, and Hume (2003) rejects the idea that there are any markedness asymmetries (at least with respect to Place of Articulation). Rice (4.7) and de Lacy (2006§1.3) re-evaluate the scope of markedness effects, arguing for recognition of a strict division between Competence and Performance. There has also been an ongoing re-evaluation of the empirical facts that support markedness. While there is much debate about which markedness asymmetries exist at the moment, it is at least clear that many traditional markedness diagnostics are not valid (e.g. Rice 4.6; also Rice 1996 et seq., de Lacy 2002a, 2006, Hume & Tserdanelis 2002).

Part I

Conceptual issues

2

The pursuit of theory

Alan Prince

2.1 The Theory is also an object of analysis

Common sense is often a poor guide to methodology. Any theory presents us with two fundamental and often difficult questions:

— What *is* it?
— How do you *do* it?

The first of these arises because a theory is the totality of its consequences. It must be given as the set of its defining conditions, and we may polish them, ground them, tailor them to meet various expectations, but unless we have mapped out what follows from them, the theory remains alien territory. Newton's theory of gravitation can be written on a postcard, and we might like to think of it as nothing more than what makes apples fall straight to earth and planets follow simple repetitive paths, but its actual content is strange beyond imagining and still under study hundreds of years after it was stated.[1] Once formulated, a theory has broken definitively with intuition and belief. We are stuck with its consequences whether we like them or not, anticipate them or not, and we must develop techniques to find them.

The second question arises because the internal logic of a theory determines what counts as a sound argument within its premises. General principles of rigor and validation apply, of course, but unless connected properly with the specific assumptions in question, the result can easily be oversight and gross error. Here's an example: in many linguistic theories developed since the 1960s, violating a constraint leads directly to ungrammaticality. A parochial onlooker might get the intuition that violation is somehow ineluctably synonymous with ill-formedness, in the nature of things. A grand conclusion may then be thought to follow:

(1) ". . . the existence of phonology in every language shows that Faithfulness [in Optimality Theory] is at best an ineffective principle that might well be done without." (Halle 1995b).

'Phonology' here means 'underlying-surface disparity'. Each faithfulness constraint forbids a certain kind of input–output disparity: *case closed*. But no version of Optimality Theory (OT) has ever been put forth that lacks a full complement of Faithfulness constraints, because their operation – their minimal violation, which includes satisfaction as a special case – is essential to the derivation of virtually every form. The intuition behind the attempted criticism, grounded in decades of experience, is that well-formed output violates no constraints; but this precept is theory-bound and no truth of logic. It just doesn't apply to OT, or to any theory of choice where constraints function as criteria of decision between flawed alternatives.

2.1.1 Optimality Theory as it is

A more telling example emerges immediately from any attempt to work within OT. At some point in the course of analyzing a given language, we have in hand a hypothesized constraint set and a set of analyses we regard as optimal. We now face the *ranking problem*: which constraint hierarchies (if any) will produce the desired optima as actual optima?

Any sophisticated problem-solver's key tactic is to identify the simplest problem that contains the elements at play, solve it, and build up from there. Let's deploy it incautiously: since the smallest possible zone of conflict involves two constraints and two candidates (one desired optimal), gather such 2×2 cases and construct the overall ranking from the results.[2] But the alert should go up: no contact has been made with any basic notion of the theory. We actually don't know with any specificity what it is about the necessities of ranking that we can learn from such a limited scheme of comparison. A wiser procedure is to scrutinize the definition of optimality and get clear about what it is that we are trying to determine. A rather different approach to the ranking problem will emerge. What, then, does 'optimal' actually mean in OT? Let us examine this question with a certain amount of care, which will not prove excessive in the end.

Optimality is composite: the judgment of hierarchy is constructed from the judgment of individual constraints. Proceeding from local to global, definition begins with the 'better than' relation over a single constraint, proceeds to 'better than' over a constraint hierarchy, and then gets optimality out of those relations.

In the familiar way, one candidate is better than another on a constraint if it is assigned fewer violations by that constraint.

(2) *'Better than' on a constraint*
For candidates a,b and constraint C, $a \succ_C b$ iff $C(a) < C(b)$.

Here we have written $a \succ_C b$ for 'a is better than b on C', and C(x) for the (nonnegative) number of violations C assigns to candidate x.

To amalgamate such individual judgments, we impose a linear ordering, a 'ranking', written », on the constraint set, giving a constraint hierarchy.

(We say C_1 *dominates* C_2 if $C_1 \gg C_2$.) Using that order, and using the definition of 'better than' on a constraint just given, we define the notion 'better than on a hierarchy'.

As usual, we will say that one candidate is better than another on a hierarchy if it is better *on the highest-ranked constraint that distinguishes the two*. (This concise formulation is due to Grimshaw 1997; a constraint is said to 'distinguish' two candidates when it assigns a different number of violations to them; that is, when one is better than the other on that constraint.)

(3) *'Better than' on a constraint hierarchy.*
For candidates a,b and constraint hierarchy H,
$a \succ_H b$ iff there is a constraint C in H that distinguishes a, b, such that
(1) $a \succ_C b$
and (2) no constraint distinguishing a and b dominates C.

To be optimal is to be the best in the candidate set, and to be the best is to have none better.

(4) *'Optimal'*
For a candidate q, a candidate set K, with $q \in K$, and a hierarchy H, q is *optimal* in K according to H, iff there is no candidate $z \in K$ such that $z \succ_H q$.

Now that we know what we're looking for, we can sensibly ask the key question: what do we learn about ranking from a comparison of two candidates (one of them a desired optimum)?

Since optimality is globally determined by the totality of such comparisons, and we are looking at just one of them, the best we can hope for is to arrive at conditions which will ensure that our desired optimum is *better than* its competitor on the hierarchy. This leads us right back to definition (3), and from it, we know that some constraint preferring the desired optimum must be the highest-ranked constraint that distinguishes them. The constraints that threaten this state of affairs are those that *dis*prefer the desired optimum: they must all be outranked by an optimum-preferring constraint. Let's call this the 'elementary ranking condition' (ERC) associated with the comparison.

(5) *Elementary ranking condition*
For q,$z \in K$, a candidate set, and S, a set of constraints, some constraint in S preferring q to z dominates all those preferring z to q.

Any constraint ranking on which candidate q betters z must satisfy the ERC. (To put it non-modally: candidate q is better than z over a ranking H of S if and only if the ranking H satisfies the ERC (5).) The ERC, then, tells us exactly what we learn from comparing two candidates.

To make use of this finding, we must first calculate each constraint's individual judgment of the comparison. A constraint measures the desired

optimum against its competitor in one of just three ways: better, worse, same. We indicate these categories as follows, writing '$q{\sim}z$' for the comparison between desired optimum q and competitor z.

(6) *Constraint C assesses the comparison q vs. z.*

Comparative relation	Violation pattern	Gloss
$C[q{\sim}z] = \mathbf{W}$	$C(q) < C(z)$	'C prefers the desired optimum'
$C[q{\sim}z] = \mathbf{L}$	$C(q) > C(z)$	'C prefers its competitor'
$C[q{\sim}z] = \mathbf{e}$	$C(q) = C(z)$	'C does not distinguish the pair'

Now consider a distribution of comparative values that could easily result from some such calculation. For illustrative purposes, imagine that the entire constraint set contains six constraints:

(7) *Typical two-candidate comparison*

	C_1	C_2	C_3	C_4	C_5	C_6
$q{\sim}z$	**L**	**e**	**W**	**W**	**e**	**L**

The relevant associated ERC declares this: C_3 *or* C_4 dominates *both* C_1 and C_6.

In any ranking of these constraints on which q is better than z, this condition must be met.

We now have the tools to examine the intuition that $2{\times}2$ comparison is the building block of ranking arguments. First, consider shrinkage of the *candidate* set. In order to narrow our focus to just 2 candidates, we exclude all the others from view. This is entirely legitimate: the hierarchical evaluation of a pair of candidates is determined entirely by the direct relation between them. Some other candidates may exist that are better than either, or worse than either, or intermediate between them, but no outsiders have any effect whatever on the head-to-head pair-internal relation. This fundamental property has been called 'contextual independence of choice' (Prince 2002b:iv), and is related to Arrow's 'irrelevance of independent alternatives' (Arrow 1951:26). It is not a truth of logic, inherent in the notion of 'comparison' or 'choice', but the premises of OT succeed in licensing it. (It is also fragile: modify those premises and it can go away, as it does in the Targeted-Constraint OT of Wilson 2001.)

Now consider the role of the constraint set, where we find no such comfort. The form of the ERC in no way privileges 2-constraint arguments: *all* L-assessing constraints must be dominated, and *some* W-assessing constraint must do the domination. If we omit an L-assessing constraint from the calculation, the resulting ERC is incomplete, and it is no longer true that any hierarchy satisfying it will necessarily yield the superiority of the desired optimum (though the converse *is* true); further conditions may be required. Leaving out C_1 from tableau (7), for example, deprives us of the crucial information that C_1 must be dominated; if it is not, then undesired z betters q.

If we happen to omit a W-assessing constraint, the associated ERC can mistakenly exclude a successful hierarchy, leading to false assertions that cannot be remedied by merely obtaining further information. This is more dangerous than L-omission when we are arguing from optimum–suboptimum pairs to the correct ranking, as when dealing with the 'ranking problem' in the course of analysis. In tableau (7), for example, we have two W-assessors, C_3 and C_4. If negligence leads us to omit C_3, say, we are tempted to the conclusion that C_4 *must* dominate C_1 and C_6. This is not sound in itself, and depending on other circumstances, it could easily turn out that C_4 lies at the bottom of the correct hierarchy, dominating nothing, with C_3 doing the work of domination demanded by (7).[3]

The logic of the theory, then, allows us to discard from any particular comparison only the neutral e-assessing constraints. Tableau (7) shrinks to 2×4, and no further. In the literature, correct handling of the ERC is not ubiquitous, and omission of constraints often rests optimistically on intuitions about relevance and likely conflict. But pairwise (or intuitively restricted) examination of constraint relations has no status. This is not a matter of convenience, taste, typography, notation, presentation, or luck. We must *do* the theory as it dictates, even in the face of common sense.

2.1.2 Using the Evaluation Metric

Let us turn to a case where reliance on intuition leads to an interesting failure to appreciate what the theory actually claims. Consider the phonological theory put forth in *The Sound Pattern of English* (*SPE*: Chomsky & Halle 1968). A vocabulary is given for representing forms and for constructing rules, which are to apply in a designated order (some cyclically) to produce outputs from lexical items. Any sample of language data, even a gigantic one, is consistent with a vast, even unbounded, number of licit grammars. Which one – note the titular definite article – is correct? It is crucial to find a formal property that distinguishes the correct grammar, if linguistic theory is to claim realism and, more specifically, if it is to address the acquisition problem, even abstractly. (It is less crucial for linguistic practice, since linguists can, and indeed must, argue for grammars on grounds of evidence unavailable to the learner.) The well-known proposal is that grammars submit to evaluation in terms of their length, which is measured in terms of the number of symbols they deploy (Chomsky 1965: 37–42; *SPE* p.334). Shorter is better, and the shortest grammar is hypothesized to be the real one. The *SPE* statement runs as follows:

(8) "The 'value' of a sequence of rules is the reciprocal of the number of symbols in its minimal representation." (*SPE* p.334, ex. (9))

Ristad (1990) has noted a potentially regrettable consequence: the highest valued sequence of rules will have no rules in it at all. We therefore make the usual emendation, left tacit (I believe) in *SPE*: that we must also take

account of the number of symbols expended in the lexicon. The length of the entire Lexicon+Rule System pairing determines the values we are comparing. A rule earns its keep by reducing the size of the lexicon.

The Evaluation Metric thus defined is entirely coherent (given a finite lexicon) and, as asserted by Chomsky & Halle, "provides a precise explication for the notion 'linguistically significant generalization'. . ." which is subject to empirical test. It seems to be the case, however, that there are literally no instances where the Evaluation Metric was put to use as defined. That is: no analysis in the entire literature justifies a proposed Lexicon+Rule System hypothesis by showing it to have the best evaluation of all those deemed possible by the theory. Is there even a case where the value was calculated?

The reason is not far to seek. Though defined globally, the metric was always interpreted locally. Typically, this was at the level of the rule:

(9) ". . . the number of symbols in a rule is inversely related to the degree of linguistically significant generalization achieved in the rule." (*SPE* p.335)

But could even be extended to rule-internal contents:

(10) ". . . the 'naturalness' of a class . . . can be measured in terms of the number of features needed to define it." (*SPE* p.400).

Of course, nothing of the sort can legitimately be asserted without building considerable bridgework between the global metric and the behavior of the local entities out of which the grammar is composed. One has the intuition, perhaps, that it can't hurt to economize locally, and therefore that one is compelled to do so. But it can easily happen in even moderately complex optimization systems that a local splurge yields a global improvement by yielding drastic simplifications elsewhere. In a highly interactive system, the results of global optimization can be all but inscrutable locally.

We can see the local–global relation playing out variously in the other examples discussed above. The idea that Faithfulness is useless when violated represents a kind of hyperlocalism focused on one candidate and one constraint; of course, nothing follows. The local relation between 2 candidates, by contrast, is preserved intact in any set of candidates that contains them, including the entire candidate set. A relation between 2 *constraints*, though, has no such local-to-global portability to the entire constraint set. What is the situation, then, with the intuitive rule-focused evaluation of *SPE* phonologies?

A question not easily answered, alas: it isn't at all clear what the 'local interpretation' might be, or how it would replace the global interpretation. To evaluate, we must compare whole grammars with different lexica, different rules, and different numbers of rules. This provides no difficulty for the global metric, which doesn't see rules or lexica at all. The local interpretation wants to compare rules, though, and so must have rules in hand and some way of finding correspondences between them across grammars to render them

comparable. This appears feasible for sets of adjacent rules, under the same lexicon, which perform identical mappings and collapse under the notational conventions; but beyond that . . . obscurity.

Stepping back from the theory, I'd suggest that the actual practice was largely based on discovering contingencies in the data, assuming that they must be reflected in rules of a specific type, and then setting out to simplify the assumed rule-types through notational collapse, ordering, and some fairly local interactional analysis; all under lexical hypotheses that sought a single underlying form for each morpheme. This is reasonable tactically, but it is a far cry from using the theory itself to compute (deterministically) which licit grammar is being evidenced by the data, and, as noted, it never involved using the theory (nondeterministically) to prove that the correct grammar had been obtained. Some such procedure of grammar discovery could even be legitimated, in principle or in part, by results clarifying the conditions under which it produces the Evaluation Metric optimum.

Overall, the effect of acting as if there were a "local interpretation" was not negative. Under its cover, attention was focused on processes, representations, their components and interactions, leading to substantive theories of great interest. Nevertheless, the divergence between theory and practice deprived the theory of the essential content that it claimed. Much effort was expended in fending off opponents who had, it seems, little knowledge of the theory they were criticizing, a faulty grasp of optimization, and little feel for how empirical consequences are derived from the actual assumptions of a theory as opposed to some general impression of them. One such defensive/offensive statement is the following:

(11) "It should be observed in this connection that although definition (9) [rephrased as (8) above] has been referred to as the 'simplicity' or 'economy criterion,' it has never been proposed or intended that the condition defines 'simplicity' or 'economy' in the very general (and still very poorly understood) sense in which these terms usually appear in the philosophy of science. The only claim that is being made here is the purely empirical one . . ."[4] (*SPE* pp.334–5)

We grant, of course, that the *SPE* theory is abstractly empirical in the way it characterizes linguistic knowledge, and note that the contemporary research style has profited enormously from the unprecedented daring exhibited in staking out territory where none before had imagined it possible. What's missing, though, is the sense of any *particular* empirical claim or set of claims which has been identified and tested against the facts. Worse, the failure to use the theory of evaluation means that we literally do not know what such a claim is. This is Newton's *Principia* without the equations, or with equations that have never been solved. Many rules and rule systems were put forth to describe many language phenomena; but in no case can we be sure that the system proposed is the one projected by the Evaluation Metric. But it is only the optimal system that contains the claims to test.

The Evaluation Metric imbroglio is directly due to a failure to apply the definition to the practice of the theory. The definition provided a formal front for the activities of the researcher, which proceeded on a separate, intuitive track. As with the example of erroneous but commonly applied beliefs about ranking, it is not satisfactory to point defensively to the success of some practitioners in developing interesting theories under false premises. "A long habit of not thinking a thing wrong, gives it a superficial appearance of being right, and raises at first a formidable outcry in defence of custom" (Paine 1776). We must do better.

2.2 What is real and what is not

One need only glance at the formal literature leading up to generative grammar to grasp that we are the beneficiaries of a fundamental change in perspective. Aiming in *Methods in Structural Linguistics* (1951) for "the reduction of linguistic methods to procedures" (p.3), Zellig Harris introduces his proposals with this modest remark:

(12) "The particular way of arranging the facts about a language which is offered here will undoubtedly prove more convenient for some languages than for others." (Harris 1951:2)

He does not intend, however, to impose a "laboratory schedule" of analytical steps that must be followed sequentially, and he characterizes the value of his methodology in this way:

(13) "The chief usefulness of the procedures listed below is therefore as a reminder in the course of the original research, and as a form for checking or presenting the results, where it may be desirable to make sure that all the information called for in these procedures has been validly obtained." (Harris 1951:1–2)

These are to be "methods which will not impose a fixed system upon various languages, yet will tell more about each language than will a mere catalogue of sounds and forms."

The goal, then, is to produce useful descriptions, to be judged by such criteria as accuracy, convenience, reliability, responsiveness to variation, and independence from observer bias. No one can sensibly dispute the importance of these factors in empirical investigation of any kind. What further ends is linguistic description intended to serve? Historical linguistics and dialect geography, phonetics and semantics, the relation of language to culture and personality, and the comparison of language structure with systems of logic are cited as areas of study that will profit from "going beyond individual descriptive linguistic facts" to "the use of complete language structures" (p.3).

Largely absent from this program is a sense that the focus of study is a real object, evidenced by the arranged facts but not reducible to them, about which one makes statements that are (because it is real) *right* or *wrong* – as opposed to convenient or awkward, useful or irrelevant to one's parochial purposes. Descriptive, synchronic linguistics is a conduit for pipelining refined information to various disciplines that make use of language data. Chomsky changes all that, of course, by identifying an object that linguistics is to be *about* – competence, I-language, the internal representation of linguistic knowledge. This move is set in the context of rival conceptions of mental structure:

(14) ". . . empiricist speculation has characteristically assumed that only the procedures and mechanisms for the acquisition of knowledge constitute an innate property of the mind. . . . On the other hand, rationalist speculation has assumed that the general form of a system of knowledge is fixed in advance as a disposition of the mind, and the function of experience is to cause this general schematic structure to be realized and more fully differentiated." (Chomsky 1965:51–52)

The ground has been shifted so fundamentally that both poles of this opposition lie outside the domain in which Harris places himself, where 'knowledge' of language is not at issue. Nevertheless, there is a clear affinity between Harris's interest in methods and the empiricist focus on 'procedures and mechanisms'. Note, too, the force of the Evaluation Metric idea in this context, since it severs the choice of grammar completely from methods and procedures of analysis: the correct grammar is defined by a formal characteristic it has, not as the result of following certain procedures.

To pursue the issue further into linguistics proper, let us distinguish heuristically between 'Theories of Data' (TODs), which produce analyses when set to work on collections of facts, and 'Free-Standing Theories' (FSTs), which are sufficiently endowed with structure that many predictions and properties can be determined from examination of the theory alone.

A near-canonical example of a TOD is provided by the Rumelhart-McClelland model of the English past tense (Rumelhart & McClelland 1986; examined in Pinker & Prince 1988). This is a connectionist network which can be trained to associate an input activation pattern with an output activation pattern. When trained on stem/past-tense pairs, it will produce, to the best of its ability, an output corresponding to the past tense of its input. No assumptions are made about morphology or phonology, regular or irregular, although a structured representational system (featural trigrams) is adopted which allows a word to be represented as a pattern of simultaneous activation. This is a fully explicit formal theory, which operates autonomously. And, once trained, a model will make clear predictions about what output is expected for a given input, whether that input has been seen before or not. It makes limited sense, however, to query it in advance of training, looking for guidance as to what the structure of human language might be; and a trained model is not really

susceptible to fine-grained analytic dissection *post hoc* either, due to the complexity of its internal causal structure. The model only takes on predictive structure when it has been exposed to data, and that predictive structure can only be investigated by presenting it with more data.

Examples of Free-Standing Theories are not difficult to find. A theory that spells out a sufficiently narrow universal repertory of structures, constraints, or processes, and explicitly delimits their interactions, will generate an analytically investigable space of possible grammars. Clear examples range from early proposals like that of Bach (1965), Stampe (1973), Donegan & Stampe (1979) to parametrized theories in syntax and those in phonology like Archangeli & Pulleyblank (1994), Halle & Vergnaud (1987), Hayes (1995), as well as many others; Optimality Theory (Prince & Smolensky 2004) falls into the Free-Standing class, both in the large and in domain-specific instantiations of constraint sets. Such theories are in no way limited to symbol-manipulation; the Dynamic Linear Model of stress and syllable structure (Goldsmith and Larson 1990, Larson 1992, Goldsmith 1994, Prince 1993), which computes with numbers, is as canonical an example of an FST as one could imagine, as we will see below in Section 3.2.

The distinction is heuristic and scalar, because theories may be more and less accessible to internal analysis, and may require more or fewer assumptions about data to yield analytical results.[5] Even a dyed-in-the-wool TOD like the Rumelhart-McClelland model admits to some analysis of its representational capacities, and Pinker & Prince mount a central argument against it in terms of its apparent incapacity to generalize to variables like 'stem' which range over lexical items regardless of phonetic content (Pinker & Prince 1988, Prince & Pinker 1988; Marcus 2001). Nevertheless, it is clear that Optimality Theory, for example, or parametrized theories of linguistic form, will admit a deeper and very much more thorough explication in terms of their internal structure.

The distinction between Theories of Data and Free-Standing Theories cross-cuts the empiricist/rational distinction that Chomsky alludes to in the passage quoted above. On the empiricist side, 'procedures and methods for the acquisition of knowledge' can be so simple as to admit of detailed analysis, like that afforded to the two-layer 'perceptron' of Rosenblatt (1958) in Minsky & Papert (1969), which treats it as an FST and achieves a sharp result. But the major step forward in connectionist theory in the 1980s is generally agreed to have been the advance from linear activation functions to differentiable nonlinear activation functions, which in one step enormously enriched the class of trainable networks and rendered their analysis far more difficult.[6] On the rationalist side, *SPE*-type phonology has a TOD character, and investigation of its fundamental properties has shown its general finite-state character (Johnson 1972) but, to my knowledge, little of research-useful specificity.

It is perhaps not surprising that many recent versions of linguistic theory developed under the realist interpretation of its goals should fall

toward the FST end of the spectrum. If the aim is to discover a 'system of knowledge' that is separate from the encounter with observables, then unless a hypothesized system has discernible properties and significant predictivity, it is unlikely to be justifiable. To the extent that it is data-dependent, and usable mostly for modeling data rather than predicting general properties, it must face off with other TODs, particularly those offering powerful mechanisms for induction and data representation. (If compressing the lexicon is the supreme goal of phonology, expect stiff competition from the manufacturers of WinZip™ and the like.) Within the ever-expanding palette of choices available to cognitive science, it seems unlikely that rationalist theory will beat statistical empiricism on its native turf. The argument must be that the object of study is not what empiricism assumes it to be. But this must be shown; and is best shown by the quality of the theories developed from rationalist assumptions.

In the absence or failure of such theories, linguistics must recede to a Harris-like position: it might serve as a helpful guide to scientists who (for whatever reason) wish to study phenomena where language plays some role, a map of the terrain but no part of the terrain itself. What's real would be the general data-analyzing methods of empiricist cognitive science, for which language has no special identity or integrity, along with whatever results such methods obtain when applied to the data, linguistic or other, that is fed to them.

In phonology proper, representational theory has moved from the undifferentiated featural medium of *SPE* to the deployment of special structures keyed to the properties of different phenomenal domains, leading naturally (though not inevitably) to contentful FSTs of those domains. Increasing the structural repertory is a two-edged sword. Poorly handled, taken as an add-on to available resources, it can turn out to be no more than a profusion of apparatus that enriches descriptive possibilities, leading to TOD. More interestingly configured, it can yield narrow, predictive theories; but these will contain significant built-in content and hence tend toward the FST side of the spectrum.

In this context, the surprise is not the emergence of the FST but the persistence of what we might call the 'Descriptive Method' (DM) – data description as the primary analytical methodology for determining the content of a theory. For a TOD, this is virtually inevitable; there may be no other way to get an inkling of the theory's character. As soon as an FST is given, though, its consequences are fully determined by its internal structure.

Yet by far the dominant approach to probing linguistic FSTs consists of confronting them with specific data. This can be done haphazardly or with reference to a few inherited 'favorite facts', or it can be done with prodigious vigor and problem-solving prowess, as in for example Hayes (1995). Although parametric theories are plentiful, few indeed are those whose 'exponential typology' of parameter settings has been laid out in full or studied in depth.

This places linguistic theory in an odd position. The axioms or defining conditions of a theory provide a starting place, not an endpoint: a theory is the totality of its consequences. With an FST, these are available to us analytically, and claims about the theory can be decided with certainty. If we decline to pursue the consequences analytically, we impose on ourselves a limited and defective sense of what the theory actually is. This then unnecessarily distorts both further development and theory comparison. Rational arguments about two theories' comparative success, for example, depend on a broad assessment of their properties; lacking that, such discussions not infrequently descend into the cherry-picking of isolated favorable and unfavorable instances.[7] What we might call the 'Analytical Method' is essential for determining the systematic content of theory. It is particularly valuable for delimiting the negative space of prohibitions into which the Descriptive Method does not venture, but it is equally essential for finding the structure of a theory's predictions of possibility.

2.3 Following the Analytical Method

Analysis of Free-Standing Theories is often driven by the most basic formal questions. Perhaps the most fundamental thing we must ask of a proposed theory is — 'does it *exist*?' That is: do the proposed defining conditions actually succeed in defining a coherent entity?[8] Closely related is the question of *under what conditions* the theory exists: what conditions are required for it to give a determinate answer or an answer that makes sense formally?[9] A natural extension of such concerns, for linguistic theories, is the question of whether the theory is *contentful* in that it excludes certain formally sensible states-of-affairs from description. It might seem to some that such questions are arid and of limited interest, since (on this view) most formal deficiencies will not show up in practice, and in the empirical hurly-burly those that do can be patched over. We have already seen how, contrary to such expectations, commanding the answers to drily fundamental questions (e.g. what *is* optimality?) is essential to the most basic acts of data-analysis. Here we examine two cases that show the very tangible value of asking the abstract questions about a theory's content and realm of existence.

2.3.1 Harmonic Ascent

Let us first consider Optimality Theory in the large. Moving beyond the barebones definition of optimality, let us endow the constraint set with some structure: a distinction between Markedness constraints, which penalize configurations in the output, and Faithfulness constraints, which each demand identity of input and output in a certain respect by penalizing any divergence from identity in that respect. Assume that the Markedness/Faithfulness distinction partitions the constraint set, so that any licit constraint belongs

to one of the categories; let's call the theory so defined 'M/F-OT'. This gives us perhaps the simplest feasible OT linguistic theory, assuming the usual generative phonological architecture in which the grammar maps a lexical form (input) to a surface form (output). We may now ask if the theory achieved at this level of generality is *contentful*, or if it requires further structure to attain predictions of interest. Exactly this question is taken up in Moreton (2004a), and the results he obtains are illuminating.[10]

To begin, we note that OT has a property that we might call 'positivity' which it shares with certain other multiple-criterion decision-making systems, though by no means all.[11] Broadly speaking, a 'positive' system will be one in which a candidate can do well globally only by doing well locally. If a *winning* candidate does poorly on some criteria in comparison to some particular competitor, we can infer, in a positive system, that it must be doing better than its competitor on some other criteria. OT's positivity comes immediately from the way it defines 'optimal': we know that if on some *hierarchy* it happens that q is better than z, then there is some particular *constraint* on which q is better than z on (namely, the highest ranked constraint that distinguishes them). Now widen the focus: suppose we know that the inferior candidate z is (perversely) better than q on some designated subset D of the constraints, ranked as in the hierarchy as a whole. Clearly, since q is the overall superior candidate, it must be that q is better than z on some particular constraint, and that constraint must belong to *the complement set* of D.

Applying this observation to M/F-OT, we find that if q, the superior candidate, is worse than z on the Faithfulness subhierarchy, then q must be better than z on the Markedness subhierarchy (and vice versa). This observation gains particular force because it is commonly the case that there is a fully faithful candidate (FFC) in the candidate set. The FFC has a tremendous advantage, because it satisfies every F constraint and nothing can beat it over the Faithfulness constraints, no matter how they are ranked. It follows that any non-faithful mapping – any mapping introducing faithfulness-penalized input–output disparity – can be optimal only if it is superior to the FFC on grounds of Markedness. Since the FFC is essentially a copy of the input, this means that in an unfaithful mapping, the output must be less marked than (the faithful copy of) the input, when it exists. We can call this property 'harmonic ascent', using the term 'harmonic' to refer to the opposite of 'markedness'.

(15) *Harmonic Ascent*
Suppose for $y \neq x$, $x \rightarrow y$ is optimal for some hierarchy H, where $x \rightarrow x$ is also a candidate.
Then for H|M, the subhierarchy of M constraints ranked as they are in H, it must be that $y \succ x$ on H|M.

Sloganeering, we can say: if things do not stay the same, they must get better (markedness-wise). See Lemma (26) of Moreton (2004a) for details.

This property severely restricts the mappings that M/F-OT can execute. A first consequence is that there can be no *circular chain shifts*. This is easiest to see in the case of the smallest possible circle: imagine a grammar that takes input /*x*/ to distinct output [*y*] and input /*y*/ to output [*x*]:

x → *y*
y → *x*

(An example would be a grammar mapping /pi/ to [pe] and /pe/ to [pi].) This pair of mappings cannot be accommodated in one grammar under M/F-OT, because the 'better than' relation is a strict order. By Harmonic Ascent, the optimality of $x \rightarrow y$ requires $y \succ x$ on the Markedness subhierarchy. But $y \rightarrow x$ requires $x \succ y$. One form cannot be both *better than* and *worse than* another.

More generally, any chain shift involving a cycle cannot be expressed. For example:

(16) *Impossible chain-shift in OT*

Mapping	Markedness Relation
$x \rightarrow y$	$y \succ x$
$y \rightarrow z$	$z \succ y$
$z \rightarrow x$	$x \succ z$

Here the argument is just one step more complicated. Putting all the implied Markedness relations together, we have $x \succ z \succ y \succ x$. Since 'better than' is transitive, asymmetric, and (hence) irreflexive, this set of relations is impossible: it yields $x \succ x$, as well as both $x \succ y$ and $y \succ x$.

A second consequence follows from this fact: there is an end to getting better. If OT is to exist at all, no constraint can portray the candidate set as an unbounded upward-tending sequence of better and better forms (see note 9). This, taken with Harmonic Ascent, rules out the endless shift:

(17) *Impossible endless shifts in OT*

$x_1 \rightarrow x_2$
$x_2 \rightarrow x_3$
$x_3 \rightarrow x_4$
…
$x_k \rightarrow x_{k+1}$
…

Of these consequences, the second seems clearly right. There is, I believe, no phonological process that, for example, adds a syllable to every input. Actual augmentation processes aim to hit some target (like bimoraicity or bisyllabicity) which is clearly relatable to Markedness constraints on prosodic structure. There is no sense in which longer is better regardless of the outcome (McCarthy & Prince 1993b, Prince & Smolensky 2004).

The first is perhaps more interesting because it characterizes rather than merely excludes. Chain shifts are well-attested, and almost always

noncircular. Moreton & Smolensky (2002) review some 35 segmental cases, of which 3 are doubtful, 4 inferred from distribution, and 28 robustly evidenced by alternations; none are circular. The famous counter-example is the 'Min tone circle' of Taiwanese (Xiamen, Amoy) tone sandhi, examined in Moreton (1999, 2004a) and much discussed in the literature (see e.g. Chen 1987, 2000, Yip 2002 and references therein). The details of the case, Moreton argues, are such that it does not invite analysis in terms of "simple, logical, plausibly innate constraints," and, as a phenomenon that is "synchronically speaking, completely arbitrary and idiosyncratic," it must be understood as a nonphonological "paradigm replacement" (Moreton 2004a:159), an intriguing possibility in need of further specification (but see Mortensen 2004 for more cases and a different view). In the end, if the circular cases prove to fall under special generalizations outside the reach of core phonology, then the prediction is vindicated. At this point, the matter must be regarded as somewhat unsettled, absent a compelling analysis of the tone circle.

Whatever the fate of circularity, it remains remarkable that a theory as simple as M/F-OT, at a level of analysis that lacks any characterization of constraints other than the formal, should show a property like Harmonic Ascent, which governs and severely restricts what it can do. We need theories that have such properties if we are to establish the rationalist perspective that Chomsky enunciated in his foundational work. The Descriptive Method of theory investigation, and its typically particularized results, can give no hint that such a property is obtainable without stipulation. Equally remarkable is the abstractness of the question that led to its discovery: 'what limitations does the theory place on the mappings a grammar can accommodate?' One might expect the answer to be so negative ('no limit') or so abstract (for example, registering them with respect to automata theory) that no obvious practical consequences ensue. Theoretically, we learn that expanding the repertory of constraint types to include *anti*-Faithfulness constraints (Alderete 1999b, 2001b) is more than an aesthetic complication; if unrestricted, it imperils the core emergent property of M/F-OT. And empirically, we find ourselves steered directly toward an entirely central phenomenon and informed that it is not merely of descriptive interest, but that its character actually determines the kind of theory we can have.

A further consequence of major analytical significance follows immediately from Moreton's work. Suppose we have a chain shift, [1] $x \rightarrow y$, [2] $y \rightarrow z$; this can only be obtained by preventing x from going all the way to z. We know from [2] that z is better than y on the Markedness subhierarchy. Thus, only Faithfulness can prevent x from leaping all the way to z; it is futile to seek a Markedness explanation for the fact that x halts at y.

More exactly, the ungrammatical candidate $^{*}x \rightarrow z$, which we wish to avoid, is better on Markedness than licit $x \rightarrow y$, but to lose, it must be *worse* on Faithfulness. This means that we need a Faithfulness constraint forbidding $^{*}x \rightarrow z$ *which does not forbid* x→y. The analysis of M/F-OT not only tells us in

general terms that circular shifts are disallowed; it specifically characterizes the kind of Faithfulness constraints that must exist if *non*circular chain shifts are to be admitted. It is far from trivial to develop a respectable theory of Faithfulness that contains such constraints; see, for example, Kirchner (1996), Gnanadesikan (1997), Moreton & Smolensky (2002), Mortensen (2004); and for other approaches, Alderete (1999b), (2001b) for antifaithfulness, and Łubowicz (2003), who aims to put the issue entirely outside the M/F distinction.

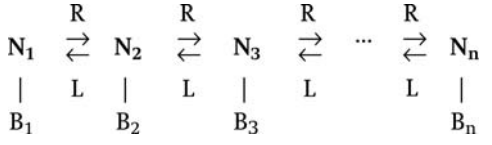
2.3.2 The Barrier Models

Goldsmith and Larson have proposed a spreading-activation account of linguistic prominence, which they have vigorously pursued through encounters with many attested patterns of stress and syllable structure — the Descriptive Method (Goldsmith & Larson 1990, Larson 1992, Goldsmith 1994). The model is, however, entirely self-contained as a formal object and susceptible to treatment as a Free-Standing Theory whose key properties can be determined analytically (Prince 1993 – henceforth IDN).[12] The aim of this section is to illustrate once again, in a very different context, how pursuing the basic formal questions leads not to an exercise in logical purification, but quite directly to properties of notable empirical significance.

The model works like this: the basic structure is a sequence of N 'nodes', each of which carries an 'activation' level, represented numerically. This gives it the power to represent ordinal properties of segments and syllables like sonority and prominence. Each node also has an unvarying bias, which may be interpreted as the intrinsic sonority or prominence of the linguistic unit that it represents. Rather than make a single calculation over these values to determine the output activation, the model calculates repeated interactions between adjacent nodes — the same mode of interaction repeated over and over. When the process converges on stable values, the model has calculated an activation profile that corresponds to a prominence structure such as a stress pattern or assignment of syllable peaks and margins. Nodes which bear greater activation than their closest neighbors – local maxima – are interpreted as having peaks of prominence.[13] Since the updating scheme is linear and iterative, we will call it the Dynamic Linear Model (DLM).

The neighborly interaction is mediated by two numerical parameters, which we designate L and R, each of which governs the character of the interaction in one of the two directions. The parameter L governs leftward spreading of activation; R, rightward spreading. Diagramatically, we can portray the situation like this:

(18) *DLM Network*



The model starts out with each node bearing zero activation. In the first step, each node gains the activation donated by its own bias; and then the serious trading begins. At each stage, the new activation of a node is determined from the current activation of its neighbors taken together with its own intrinsic bias level. The update scheme, in which we write $\mathbf{a_k}$ for the activation of N_k, can be represented like this:

(19) $\mathbf{a_k} \leftarrow \frac{1}{2}\,\mathbf{L}\cdot\mathbf{a_{k+1}} + \frac{1}{2}\,\mathbf{R}\cdot\mathbf{a_{k-1}} + \mathbf{B_k}$

A node's own current activation plays no role in determining its next state: only its bias, which never changes. Since L, R, and B_k are all constants, this is a linear scheme: each node's new activation is a weighted sum of its neighbor's activations, with its own bias added in.

Here are some examples to give a sense of how it works. Suppose we start out with a bias sequence (1,1,1,1,1,1), representing a string of 6 undifferentiated syllables. Let L=R= −1. The result is approximately (1.1, −0.3, 1.4, −0.6, 1.7, −0.9). This may look like nothing more than a mess of numbers, but the significant fact is the location of the local maxima – those nodes greater than their neighbors (or neighbor, if at an edge). Marking those, we see that the DLM has calculated this mapping, which we write using x for 'unstressed' and X for 'stressed': x x x x x x → X x X x X x

A familiar kind of alternating pattern has been imposed.

Now suppose we start out with a bias sequence (0,0,1,0,0,0) and set L=1.333 and R=.75. The result comes out approximately like this: (2.0, 3.0, **3.4**, 1.9, 1.0, 0.4). Identifying the one maximum (bolded), we see that this is the Input → Output relation:

x x X x x x → x x X x x x

which is naturally interpreted to express a case in which an accent marked in the lexical input has been preserved on the surface.

If we alter the L,R parameters, we get a different result: for L=1.6, R=.635, we get approximately (2.9, **3.7**, 3.4, 1.6, 0.7, 0.2). The significant configuration now centers on the second entry, and we have portrayed the map

x x X x x x→ x X x x x x

in which an underlying accent has been over-ridden.

A variety of linguistic and nonlinguistic patterns may be produced from such experimentation, suggesting the value of further systematic research.[14] What, then, are the general properties of the theory? At this point, two paths diverge. We may follow the Descriptive Method, with Goldsmith and Larson, aiming to deal with a wide range of known prominence phenomena in specific languages by finding L, R values and biases that will accommodate them. Or we may attempt to see what we can learn by interrogating the formal structure of theory, trying to classify its parameter space and look for characterizing properties.[15]

Let's start with one of the most fundamental questions we can ask: under what conditions does the theory *exist*? In the context of an iterative scheme

like the DLM, this question takes a clear and exact form: when does the model converge, producing stable finite values as output? Specifically, what values of the parameters L and R lead to convergence? The fine-grained convergence limit is tied to a specific model's length in nodes; but generalizing over all models, we have this pleasing result, which will prove quite useful: if the absolute (unsigned) value of the product L·R is less than or equal to 1, any model of any length will converge.

(20) *Convergence of the DLM*
Any Dynamic Linear Model M_n with $|LR| \leq 1$ converges, for all n, n the number of nodes in the model.[16] (IDN:53)

From the descriptive point of view, this result has its uses – it tells us where not to look for parameter values – though, in practical terms, if we start our search near zero for both L and R, an astute prospector armed with a spreadsheet program ought to be able to find suitable values experimentally, when they exist. Analytically, its interest emerges when we ask a further question, targeted at finding the content of the theory in its realm of existence: given L, R, and a sequence of biases, is there a *formula* that describes the output of the iterative scheme? The goal is not merely to shorten the process of calculation (pointless in the Excel™ era), but to have a characterization of the model's output that may be scrutinized for general properties.

For the vast majority of networks, 'solving the model' in this way is not an option, and the Descriptive Method is essential to finding out what's going on; this is why we classified the Rumelhart & McClelland model as a TOD, and why people tend to think of network models as TOD on arrival. But the simple structure of the DLM renders it amenable to analysis.

Because the function computed by the DLM is linear in the biases, it is natural formally to inquire about the fate of bias sequences that consist entirely of 0's except for a single 1. Any other sequence can be built up from a weighted sum of such basic sequences. Here linguistics lines up happily with algebra – it is also linguistically natural to regard such sequences as representing a form with a single lexical accent.

We want to describe the value assumed by each node, given that the 'underlying accent' occurs in a certain place. The local maximum in the output, which is fully determined by these values, is where the surface accent lies. Calculation produces a formula which is a bit messy though not intractable (involving hyperbolic sines and cosines and the occasional complex number; see IDN:62). But a remarkable simplification occurs when we restrict the parameters to the curves LR=1, on which convergence is universally guaranteed.[17] Because of their simplicity, we may call these the 'Canonical Models'. The Canonical Models come in two kinds. Either L and R are both negative, in which case we have alternation of prominence, as we always do when both parameters are negative; or both parameters are positive.

The behavior of the general DLM when both L and R are positive is straightforward: accent is culminative, with a single maximum occurring in the activation function.[18] The same will be true in the Canonical Models. But when we seek the location of that maximum in the Canonical Models, a striking property emerges: there is a *window* at one edge or the other into which the surface accent must fall.

Given any value of R greater than 1, the surface accent can fall no further than a certain distance from the right edge, regardless of where the underlying accent is placed. The same is true for L (corresponding to values of R less than 1), with respect to the beginning of the word. Within the window, underlying accent is preserved. Outside the window, it is lost and in its place, as it were, the accent shows up at the inner edge of the window – the closest unit to the underlying accent that can be surface-accented.

We can name each model by the farthest internal location at which an accent can fall, (given single accented input), indicating by subscript the edge it measures from: thus, 3-Model$_R$ is the model in which the accent can fall no further into the string than the 3rd node from the end. Let us call these Canonical Models the 'barrier models', since in a k-Model, the kth node provides a kind of barrier beyond which surface accent may not venture. The parameter space divides up as in Table (21). NB: the cited ranges *exclude* the end points.

(21) *Right Barrier Models*

Model #	"range" of R	Length of Range	Accent no further from end than
1-Model$_R$	∞ to 2	∞	final syllable
2-Model$_R$	2 to 3/2	1/2	penult
3-Model$_R$	3/2 to 4/3	1/6	antepenult
4-Model$_R$	4/3 to 5/4	1/12	preantepenult
5-Model$_R$	5/4 to 6/5	1/20	prepreantepenult
...	...	...	
j-Model$_R$	j/(j–1) to (j+1)/j	1/j(j–1)	(pre)$^{j-3}$ antepenult

Symmetrically, the Left Barrier Models determine a window at the *beginning* of the string. The Right Barrier Models charted above occupy the parameter span where R $\in$ (1, ∞). The Left Barrier Models lie within the positive line segment L $\in$ (1, ∞), or equivalently R $\in$ (0,1), since R=1/L.[19]

This result is multiply remarkable. First, the barrier/windowing behavior is fully emergent from assumptions which make no mention of anything like that property. The alternating pattern that comes about when L and R are both negative has a kind of resonance with structural formulations like *Clash (Kager 9.2.1). Both, in their different ways, seek to suppress prominence on adjacent units. And when L and R are both positive, it is perhaps not naively expected that the result should be a single maximum

in the activation function, but it doesn't seem like an unusual outcome. It is the particularity of the windowing effect, and its lack of reducibility to some obvious local characteristic of the network, that makes it surprising.

Second, it is remarkable that the parameter ranges are valid for any length of string.[20] The number of nodes plays a role in the formula describing the output, and in other situations it figures in empirically anomalous dependencies (IDN:17). In this case, though, we have conditions that are valid across all forms, fully independent of form size.

Third, although nontrivial barrier/windowing behavior, with non-peripheral accents allowed, goes on outside the Canonical Models, it is restricted to a relatively small portion, a little less than 1/6, of the parameter space in the first quadrant. This means that random prospecting could easily miss it. Crucial to finding it is investigation along the hyperbola LR=1; but this curve presents itself as particularly interesting only because of its role in delimiting convergence.[21] The abstract, airless-seeming question with which we began – under what conditions does the model exist? – has led us right to one of its central properties.

Finally, it is striking that this fundamental result connects directly with a major phenomenon in stress and accent systems. The DLM overshoots the mark in a couple of respects – it is totally left-right symmetric, and allows windows of any size, while known windowing systems typically range up to no more than 3 syllables in length at the end of words, and 2 syllables at the beginning.[22] Whatever the remaining questions, the model opens the way to an entirely novel account of the windowing effect, unlike anything seen before. This renders the DLM worth studying alongside the other contentful accounts of prosodic structure that occupy linguistic attention, while vindicating the analytic method that reveals its structure.

2.4 Description and descriptivism

In a recent essay, Larry Hyman asks and answers the question "Why Describe African Languages?" (Hyman 2004). He argues that there is irreducible value in describing "complex phenomena using the ordinary tools of general linguistics," and that this goal stands in opposition to, and is at least as worthy as, developing grammars within current "theories [that] are not description-friendly," such as Minimalism and OT.

With the main thrust of his argument there can be little dissent: deep empirical work discovering the facts and generalizations of human languages is the very basis of linguistics, and it is essential that there be sound descriptions to convey them to the community of researchers. Why then the question? In part, Hyman's concern is driven by disciplinary attitudes toward 'theory' and 'description' – where, it seems, a certain class of person expects one to make a 'theoretical contribution' in every outing and will disdain or suppress work that lacks that key ingredient.[23] As for

what a 'theoretical contribution' might be, Hyman cites an unidentified commentator:

(22) "The shared belief of many in the field appears to be that a paper making a theoretical contribution must (a) propose some new mechanism, which adds to or replaces part of some current theory, or (b) contradicts some current theory. Papers that do neither, or those that do either but in a relatively minor way, are not looked at as making a theoretical contribution." Quoted in Hyman (2004:25).

This is very much a matter of 'mind your labels' – and we shouldn't be led to abandon the idea of 'theoretical contribution' because an obtunded version is instrumental in the intercollegial jostling and jousting of the field. In the present context, where a theory is taken to be an object in grave need of explication and analysis, it should be clear that an authentic 'theoretical contribution' can involve deepening the understanding of a theory's consequences or of the proper methods of using it, without a hint of replacement or contradiction.[24] We reject the 'shared belief' identified in the quote, and deny the privileged status it accords to certain types of work, to advocate a broader though not boundaryless account of what a contribution, including a 'theoretical contribution', may be. Hyman's move, by contrast, is to argue toward a unification of theory with description, neutralizing the distinction: "description and theory are very hard to disentangle – and when done right, they have the same concerns" (p.25). He goes on to clarify:

(23) "Description is *analysis* and should ideally be

(a) rigorous . . .
(b) comprehensive . . .
(c) rich . . .
(d) insightful . . .
(e) interesting . . ." (Hyman 2004:25)

No one would dispute either the importance of the cited criteria or the claim that they apply to theory as well as description. A closer look, though, is profitable, and suggests some important divergences. Criteria (c), (d), and (e) are contentful but difficult to assess intersubjectively, and perhaps connect more closely with Harris's 'convenience' than with questions of truth and falsity. We therefore focus on (a) *rigor* and (b) *comprehensiveness*.

Of rigor, the key remark is the one made in Section 1 above: there is no general sense of rigor that can be directly applied without regard for the specific assumptions at play in a given case. Work is therefore required. To design a successful ranking argument, as in our example, you must build from the actual definition of 'optimality'. It is necessary to ask 'what can be learned from the comparison of two candidates, one assumed optimal?' If the Evaluation Metric is to be employed seriously, you must inquire about the relation between local reduction of symbol consumption and the

eventual global symbol count of the entire grammar. To achieve 'rigor', there is a range of questions that must be asked about the theory itself, and these questions differ in character from those asked of data (e.g. what is the distribution of downstepped high tone in Bangangte Bamileke?) or of the data-analysis relation (e.g. how are floating tones interpreted? how are they manipulated in Bangangte Bamileke?).[25] And different methods are required to answer them.[26]

Comprehensiveness – the inclusion of all relevant material – is a systematic notion and therefore presupposes a notion of 'system' which delimits relevance. Just like rigor, then, it takes on different colorations in different contexts. Contrast the questions to be asked and the techniques required to attain and evaluate, say, a full account of a language's verbal paradigm[27] with those used to derive and characterize the consequences of a formal theory. It makes sense to classify these as different 'contributions', if we are classifying things, though the inevitable ensuing scuffle to hierarchize them socially is better explicated by primatology than by the philosophy of science.

In the present context, the interpretation of *comprehensiveness* also marks an important divide between appropriate strategies for descriptive work and for theory development. Much can be gained theoretically by explicitly failing to be comprehensive over the data in ways that would be absurd descriptively. The study of idealized, delimited problems is a familiar and essential tool for exploring theories. At the grand level: the de Sitter cosmology imagines a universe that lacks matter entirely (it expands); Schwarzschild solves the field equations of General Relativity under the assumption of strict spherical symmetry of matter distribution (local collapse can result).[28] To cite a case considerably humbler and closer to home: much can be learned by working with a simplified Jakobsonian typology of syllable structure (Clements & Keyser 1983, Prince & Smolensky 2004), although it would be grossly inappropriate to claim comprehensiveness for a *description* of natural language syllable patterns that overlooks long vowels, diphthongs, and intrasyllabic consonant clusters.

Investigation of theories, even via the Descriptive Method, is tied to the availability of research strategies that idealize and delimit, deferring comprehensiveness. In the case of FST, this is particularly crucial because it opens up possibilities for obtaining analytical results when the general situation is complex and its structure obscure. Attitudes toward comprehensiveness therefore play a subtle but central role in estimating the relative promise of different research directions. One line of thinking finds expression in "Why Phonology is Different" (Bromberger and Halle 1989). The authors are concerned to justify their belief that phonology is intrinsically not amenable to being understood as the interaction of universal principles, distinguishing it in their view from syntax; the key, they argue, is the availability of stipulated language-specific rule-ordering in phonology alone:

(24) "Rule ordering is one of the most powerful tools of phonological description, and there are numerous instances in the literature where the ordering of rules is used to account for phonetic effects of great complexity." (Bromberger & Halle 1989: 59).

The perspective here is determinedly descriptive; the theory is to be justified by its ability to portray "complex" cases, for which much "power" is thought to be needed. There is no hint of an ambition to find and derive general properties of the language faculty, and consequently no willingness to tolerate the local costs of such ambition — idealization; plurality of theoretical lines; openness to ideas that limit rather than expand descriptive options; empirical lacunae and anomalies; admission of uncertainty. Their argument continues:

(25) "Until and unless these accounts are refuted and are replaced by better-confirmed ones, we must presume that Principle (7) [extrinsic ordering – AP] is correct." (Bromberger & Halle 1989:59).

One can only admire the authors' willingness to take on the entire literature in an area before rejecting its premises, but there are sound reasons why this strategy has never had much purchase on the field, which has been more notable for innovation than uniformity. At bottom, providing unsteady foundations, is an unexamined notion of 'confirmation', without which such qualifiers as 'better-confirmed' and 'correct' risk vacuity. More concretely, there are so many active, promising lines of investigation into every aspect of the enterprise, from the nature of the data to the identity of the targets of explanation, that it seems premature to shut them down on the basis of a presumption.

Whatever the ultimate status of their imperative, its interest in the present context is its orthogonality to the kind of theoretical concerns we have been probing. There is no sense in their work that a theory is an opaque object, whose content and proper handling must be discovered before we can declare success and failure, even descriptively, or compare it properly with other theories. Supreme is the goal of 'accounting for', and given a disposition to regard the facts as a fixed body, the approach merges with classic descriptivism. The real threat to their favored theory, then, is not provided by those versions of generative phonology which pursue very different explanatory goals, but rather by statistical empiricism, which also avails itself of 'powerful tools' to gain even more comprehensive models of their data.

2.5 Conclusion

The encounter with fact is essential to the validation, falsification, and discovery of theories. But as soon as a theory comes into existence, it must also be encountered on its own terms. A theory cannot even be faced with fact – we cannot *do* it properly – if we don't know how to construct valid

arguments from its premises. And since a theory's content is the set of its consequences, which are typically far from legible in its defining conditions, we are obliged to interrogate its structure to find out what it *is*. Asking the fundamental formal questions, and finding or developing techniques to answer them, is an irreplaceable aspect of linguistic research that identifies the major predictions and particularly meaningful empirical challenges associated with a theory.

Linguistic theory has shown a notable tendency to develop what we have called Free-Standing Theories, those which have an internal structure susceptible to detailed analysis independent of the factual encounter. The reasons for doing so may be, as suggested above, intrinsic to the realist project, since rationalist theories require an abstract object of study whose existence is likely to be justifiable only in terms of deep, non-obvious properties. In the absence of such properties, empiricist inductivism exerts a strong claim to the territory.

It is reasonable to ask, then, why the 'Analytic Method' of confronting theories on their own terms does not play a more conspicuous role in the current ecology of the field, which could be argued to conserve, largely, an intuitive methodology more properly rooted in the descriptive ambitions of pre-generative work. An important factor may be the sense that formal analysis can be successfully replaced by approaches more closely allied to facts and to techniques for dealing with facts – 'the ordinary tools of general linguistics'. Invaluable in empirical assessment of claims, the Descriptive Method has often been taken as the primary mode of exploring a theory's structure and content, where it has severe limitations. Adhered to strictly, it cannot distinguish between a superset theory ("too powerful") and a proper subset theory; it has no particular relation to a theory's systematic properties; and it is unable to provide certainty in the assessment of claims about predictions and exclusions.

A more recent development which is sometimes taken to provide a feasible substitute for analysis is 'grounding' – in the case of phonology, pointing to phonetics as supporting the correctness of theoretical assertions. In much work, the term has a specific well-defined sense which gives it theoretical status (Archangeli & Pulleyblank 1994, Hayes 2004a:299), but it also leads a second, more fluid life as a motivator and recipient of intuitive appeals. Some of this may be discerned in the following statement from Hayes (2004a:291), who is asking "what qualifies a constraint as an authentic markedness principle?":

(26) "The currently most popular answer, I think, relies on typological evidence: a valid constraint 'does work' in many languages, and does it in different ways.

However, a constraint could also be justified on functional grounds. In the case of phonetic functionalism, a well-motivated phonological constraint would be one that either renders speech

> easier to articulate or renders contrasting forms easier to distinguish perceptually. From the functionalist point of view, such constraints are a priori plausible, under the reasonable hypothesis that language is a biological system that is designed to perform its job well and efficiently." (Hayes 2004a:291).

But the symmetry is illusory. A constraint, in the intended sense, is a principle within a theory and, like any other principle in any other theory, is justified by its contribution to the consequences of that theory. Since OT is a theory of grammar, the consequences are displayed in the grammars predicted and disallowed – 'typological evidence'. A constraint which cannot be justified on those grounds cannot be justified. Further, 'justifying' a constraint functionally (or in any other extrinsic way) can have no effect whatever on its role within the theory. A constraint, viewed locally, can appear wonderfully concordant with some function, but this cannot supplant the theory's logic or compel the global outcome ('efficiency') that is imagined to follow from the constraint's presence, or even make it more likely.

A ranking argument based on two candidates, one desired optimal, remains valid whether the constraints are grounded or not; and in Targeted Constraint OT, where grounding is invoked to support the notion of targeting (Wilson 2001:156–160), such two-candidate arguments lose their validity because of the formal structure of the theory, and phonetic function cannot restore it. The property of Harmonic Ascent cannot be abrogated, amended, or influenced by grounding or its lack. The choice of *Markedness* constraints, no matter how grounded, cannot by itself predict grammatical behavior, because mappings are determined by the interaction of Markedness with Faithfulness constraints, whose properties are crucial to the range of possible outcomes.

When stated explicitly (p.299), Hayes's 'inductive grounding' is not an exercise in the plausible,[29] but a concrete proposal for the generation of certain kinds of constraints from specific data, which relies on finding the local maxima in a certain space of possibilities. Its fate is in the hands of geometry and logic. As an actual theory, it has left behind any hopes that attended its conception and birth, and now lives in the realm of the issues explored here.

Such considerations suggest a bright future for linguistic research as it grows beyond its origins. Analysis is deaf to our desires, but it can tell us what we want to know, if we know how to ask.

Notes

I'd like to thank Paul Smolensky, John McCarthy, Jane Grimshaw, Bruce Tesar, Jean-Roger Vergnaud, Vieri Samek-Lodovici, Chaim Tannenbaum, Seth Cable, Naz Merchant, and Adrian Brasoveanu for interactions which

have shaped and re-shaped my views on the matters addressed here. Thanks to Paul de Lacy for valuable comments on an earlier draft.

1 Saari (2005) is a recent study. To get a sense of what can happen, see Ekeland (1988), esp. pp. 123–131.

2 The intuition gets a boost from previous analytical practice: in ordering rules, the analyst typically looked at two rules at a time (and that worked, didn't it?).

3 If an erroneously truncated ERC has excluded the correct hierarchy, there will be further information that contradicts it, yielding the impression that no correct hierarchy exists. Even if the erroneous ranking condition has not excluded the correct hierarchy, it produces a distorted account of the explanatory force of the various constraint relations in it.

4 Interestingly, the actual on-the-ground interpretation of the Evaluation Metric may have been closer to the loose general sense of 'be simple' than to the formal definition of evaluation.

5 At a considerably more abstract level, there is much to be said about the capacities and dynamics of connectionist networks, see Smolensky et al. (1996) for a large-scale multi-perspective overview.

6 See Rumelhart & McClelland (1986), McClelland et al. (1986a). The general view taken there is that "the objects referred to in macrostructural [i.e. symbolic –AP] models of cognitive processing are seen as approximate descriptions of emergent properties of the microstructure" (McClelland, Rumelhart, and Hinton 1986:12). Smolensky and Legendre (2005) develop a very different view, according exact reality to both continuous (micro) and discrete (macro) processing as distinct levels.

7 Interestingly, competition often provokes localized analysis of a rival theory, treated as an FST, even in the context where the favored theory is being laid out and investigated by the Descriptive Method. To cite merely one example: in Halle and Vergnaud (1987), an important synthetic work that brings together much prior theory under the unifying rubric of the bracketed grid (Hammond 1984), there is an argument against one of Hammond's proposals, based on an apparently false consequence derived from it (p.75). Halle & Vergnaud's system is well and even elegantly formalized, yet due to their reliance on the Descriptive Method, we have little idea of the scope of their own predictions, some of which may involve equally disturbing pathologies.

8 Nonexistence isn't the worst thing that can happen. Yang-Mills theory, for example, is said to be basic to modern particle physics, but is not known to 'exist' mathematically, i.e. to have coherent foundations. The Clay Institute offers $1,000,000 for showing its 'existence': http://www.claymath.org/millennium/Yang-Mills_Theory.

9 For example, the theory of multiplication and division exists; but you can't divide by zero. Similarly, if you are computing probabilities, they must not be less than 0 or greater than 1. To move nearer to our concerns, note that it is crucial for OT that there be at least one *best* element in the candidate

set. Suppose that a constraint was posited to offer *rewards* rather than penalties, as all do now. Let the putative constraint LONG give a reward of +1 for each syllable that a form contains. Then there is no candidate that has the maximal value on LONG, and were the constraint asked to produce the class of forms that do maximally well on it, no output would be defined. If such a constraint is admitted, the theory ceases to exist.

10 The presentation of Moreton's results given here will be considerably more qualitative than Moreton's own, and will diverge in some points of perspective. See Moreton (2004a) for a scrupulous rendering of the details.

11 'By no means all'—this innocuous phrase hides the difficulty, in many circumstances where ordinal preference is involved, of finding a system that has the property. Common sense intuition fails dramatically here. See Saari (2001), for example, to make contact with the vast literature emerging from Arrow (1951).

12 Discussion is based on "In defense of the number *i*" (Prince 1993 – IDN), improved notationally and formally in a few respects.

13 Although the model operates internally on numbers, it does not strive to compute an empirically-determined numerical value; its interpreted output is fully discrete and indeed binary, discriminating only peaks from nonpeaks.

14 Such experimentation with the parameters of a theory is a part of what we are calling the Analytic Method, though here we are emphasizing the aspects of analysis that yield provable results.

15 In noting this methodological divergence, we are of course not asserting that only one path should be pursued.

16 For a specific length N, we have convergence iff $|LR| < 1/\cos^2(\pi/(N+1))$, which is always greater than 1. If L and R have the same sign, a model diverges to infinity at and beyond the limiting value; if they have different signs, the model enters an oscillatory regime of period 4 at the limiting value, and diverges to infinity beyond it.

17 The resulting formula turns out to involve the product of two linear terms, each reflecting distance to the edge, and an exponential term based on either of the L or R parameters, whose exponent reflects the distance between the underlying accent and the node whose value is being computed. Schematically, we can write it like this, using $\mathbf{a}_k[j]$ to mean the value of the j*th* node in the output vector whose input has a '1' in position k and zeroes elsewhere:

$$a_k[j] = C \cdot \text{dist-k\#}(j) \cdot \text{dist-j\#}(k) \cdot R^{\text{dist}(j,k)}$$

where C is a length-based constant $2/(n+1)$, the 'tilt' $\sqrt{(R/L)} = R$, dist-k#(j) gives the unsigned distance of j from the edge where k is not in the j-to-edge path, dist-j#(k) *mutatis mutandis*; dist(j,k) is the signed distance (j – k) between j and k.

18 Caveat: what we are calling a 'maximum' can be spread across two adjacent nodes that have identical activation values.