Language and Linguistics Compass 1/4 (2007): 260–291, 10.1111/j.1749-818X.2007.00018.x

What Is Optimality Theory?[1]

John J. McCarthy*
University of Massachusetts Amherst

Abstract

Optimality Theory is a general model of how grammars are structured. This article surveys the motivations for Optimality Theory, its core principles, and the basics of analysis. It also addresses some frequently asked questions about this theory and offers suggestions for further reading.

1. Introduction

In 1991, Alan Prince and Paul Smolensky began presenting their work on a new approach to language. By 1993, this new approach had a name – Optimality Theory (OT) – and it became known through their widely circulated manuscript 'Optimality Theory: Constraint Interaction in Generative Grammar' (Prince and Smolensky 1993/2004). The impact of this work on the field of phonology was extensive and immediate; since 1993, it has also stimulated important research in syntax, semantics, sociolinguistics, historical linguistics, and other areas. Optimality Theory would probably go on anyone's list of the top five developments in the history of generative grammar.

This survey article sketches the motivations for OT, its core principles, and the basics of analysis. It also addresses some frequently asked questions about this theory. Because any survey of such a broad topic is necessarily incomplete, each section concludes with suggestions for further reading.

2. The Motivations for OT

Since the early 1970s, it has been clear that phonological and syntactic processes are influenced by constraints on the output of the grammar. Two different kinds of influence can be identified:

(i) Processes can be **blocked** by output constraints. In Yawelmani Yokuts (now usually called Yowlumne), syllables are maximally consonant–vowel–consonant (CVC). A process of final vowel deletion is blocked from applying when it would produce an unsyllabified consonant (Kisseberth 1970): /taxaː-kˀ**a**/ → [ta.xakˀ] 'bring!' vs. /xat-kˀ**a**/ → [xat.kˀ**a**], *[xat.*kˀ*] 'eat!' (the period/full stop marks syllable boundaries. Unsyllabified consonants are in italics). In English, *wh* movement is blocked when it would create a *that*-trace sequence (Chomsky and Lasnik 1977): *Who did you say* **t** *left?* vs. **Who did you say* ***that*** **t** *left?*

(ii) Processes can be **triggered** by output constraints. In Yawelmani, unsyllabifiable consonants in the input are fixed by epenthesizing a vowel: /ʔilk-hin/ → [ʔi.l**i**k.hin], *[ʔil.*k*.hin] 'sings'. In Spanish, clitics are moved to satisfy an output constraint requiring that second person pronouns precede first person, regardless of whether they refer to the direct object (DO) or indirect object (IO) (Perlmutter 1971): ***Te***$_{IO}$ ***me***$_{DO}$ *presento*/***Te***$_{DO}$ ***me***$_{IO}$ *presentas* 'I introduce myself to you'/'You introduce yourself to me'.

Although the blocking and triggering relationships between the processes and the output constraints are easy to grasp at an intuitive level, it is not obvious how to express these relationships formally in linguistic theory. In fact, phonological and syntactic theory, which developed along parallel lines until the 1970s, diverged on exactly this point.

The mainstream syntactic approach to blocking and triggering is represented by works like Chomsky and Lasnik (1977). All transformations are optional. An input to the grammar – that is, a deep structure – freely undergoes any, all, or none of the transformations, and the result is submitted to the surface-structure constraints, called filters. Any candidate surface structure that satisfies all of the filters is a well-formed sentence of the language. A transformation T is in effect blocked whenever a surface filter rules out the result of applying T. Similarly, T is in effect triggered whenever a filter rules out the result of not applying T. The net result is that much of the burden of explaining syntactic patterns is shifted from the theory of transformations to the theory of filters, which is arguably better equipped to produce such explanations. In Government-Binding (GB) theory, the theory of transformations ultimately withered away to just Move α (Chomsky 1981).

The filters model (hereafter FM) was extremely influential in syntax, but not in phonology. The main reason, in my view, is that FM required all filters to be inviolable, and this assumption will not work with typical phonological data. (It does not work in syntax either, but that limitation was not so immediately apparent.)

Yawelmani can be used to illustrate this. In FM, as we saw, obligatory application of a process is obtained by positing an output constraint against structures where that process has not applied. In Yawelmani, epenthesis is obligatory before consonants that cannot be syllabified within Yawelmani's restricted CVC syllable canon: /ʔilk-hin/ → [ʔil**i**khin]. In FM, this means that the epenthesis transformation is optional, but failure to epenthesize is marked as ungrammatical by a constraint against unsyllabified consonants in the output: *[ʔil.*k*.hin]. Final vowel deletion is another obligatory process, so some output constraint must be marking *[ta.xaː.kʔ**a**] as ungrammatical. An obvious candidate for this constraint is a general prohibition on final vowels, *V#. But this cannot be correct because it also marks as ungrammatical those well-formed surface structures where deletion has correctly failed to apply, such as /xat-kʔ**a**/ → [xat.kʔ**a**]. So the constraint responsible for the obligatoriness of deletion in /taxaː-kʔ**a**/ → [ta.xakʔ] has to be something more specific: a prohibition on final vowels that are preceded by a single consonant, *VCV#.

This seemingly innocent analytic move disguises a fundamental failure of explanation. The constraint *VCV# stipulates something that should be accounted for independently: final vowel deletion is blocked in a VCCV# context because then it would produce an unsyllabified consonant: /xat-kʔ**a**/ → *[xat.*kʔ*]. With inviolable constraints there is no way of saying that *V# is enforced unless enforcing it would produce an unsyllabifiable consonant. There is no way of establishing such priority relationships between two inviolable constraints; at least one of the constraints (the lower priority one) has to be violable for priority relationships to make any sense.

Another reason for FM's lack of influence in phonology is the continuing force of the phonological research program defined by The Sound Pattern of English (SPE) (Chomsky and Halle 1968). This program strongly influenced the kinds of explanations that could be entertained. SPE's central hypothesis, embodied in the evaluation metric, is that rules statable with a few features have greater explanatory value than rules requiring more features, ceteris paribus. The SPE theory supplies abbreviatory conventions that capture generalizations by allowing certain rules to be stated in a more compact form. Kisseberth (1970) proposed a theory of blocking effects along these lines. The idea is that the output constraint against final clusters can be used to abbreviate the rule of final vowel deletion, simplifying it from V → Ø / VC___# to just V → Ø / ___#. The problem with this proposal is that it does not extend to the triggering side of conspiracies. Yawelmani's epenthesis rule Ø → i/C___CC could not be abbreviated as just Ø → i, because the theory lacked any sort of economy mechanism to ensure that Ø → i applies only when it is needed and not otherwise.

Novel theories of phonological representation, which rose to prominence beginning in the mid-1970s, led to progress of a different sort on the blocking and triggering problems. As phonological representations became increasingly complex, it became possible to imagine an almost rule-less phonology in which automatic satisfaction of universal constraints on representations was all that mattered. Goldsmith (1976a,b) and Prince (1983) developed proposals along these lines for autosegmental and metrical phonology, respectively. This work ran headlong into another problem, however: the proposed universal constraints did not hold in every language all of the time. That is why the subsequent literature on autosegmental and metrical phonology, such as Pulleyblank (1986) and Hayes (1995), places much more reliance on rules than constraints.

This last point brings us to another of the primary motivations for OT: the problem of constraint universality or the lack of it. In both phonology and syntax, initially plausible candidates for universal constraints often become increasingly dubious as they are surrounded with an apparatus of hedges and parameters to deal with empirical challenges.

The constraint requiring all syllables to have initial consonants [Ito's (1989: 222) Onset Principle] supplies a nice example of how a good constraint can go bad if parametrization is taken seriously. In Arabic or German, every syllable must have an onset, *tout court*, and [ʔ] is epenthesized whenever it is needed to ensure that outcome. In Timugon Murut (Prentice 1971), on the other hand, onsetless syllables are tolerated, so the Onset Principle needs to be parametrized: [onsets required: yes/no]. In Axininca Campa (Payne 1981), onsetless syllables are tolerated word initially, but not word medially, so a further parameter is required: [onsets required medially: yes/no]. Dutch is similar to German, except that it allows syllables to be onsetless if they are unstressed (Booij 1995: 65), so yet another parameter will be necessary. Of course, no analyst has seriously proposed such nuanced parametrization of this constraint, but the facts would seem to demand it of anyone with a serious commitment to a parametric theory of phonological typology.

Even such complex parametrization is doomed to fail, however, because constraints can still be active in languages that seem to violate them freely. Timugon Murut, as was just mentioned, allows onsetless syllables: [am.bi.lu.o] 'soul'. Onsetless syllables are nevertheless avoided in two circumstances. A single intervocalic consonant is syllabified as an onset, such as the [l] of [am.bi.lu.o], not *[am.bil.u.o]. And reduplicated syllables must have onsets, even at the expense of infixing the reduplicative morpheme (in boldface) by copying the second syllable instead of the first one: [a.ba.lan] 'bathes' reduplicates as [a.**ba**.ba.lan], not *[**a**.a.ba.lan] (with consonant-initial words, infixation is not necessary to get an onset, so [bu.lud] 'hill' reduplicates as [**bu**.bu.lud]). Examples like these show that a constraint can remain 'on' in the phonology of a language even when it would seem to have been turned off parametrically.

By the end of the 1980s, there was certainly a consensus about the importance of output constraints, but there were also major unresolved questions about the nature and activity of these constraints. That 'conceptual crisis at the center of phonological thought', as Prince and Smolensky (1993/2004: 2) refer to it, was not very widely acknowledged at the time, but in hindsight it is hard to miss. It is a major feature of the intellectual context in which OT was developed.

Suggestions for further reading. McCarthy (2002: 48–65) is a short overview of the intellectual context of OT. Some of the main issues that arose in the field of phonology in the years after the publication of SPE are given a historical perspective in Anderson (1985: 328–50). Prince and Smolensky (1993/2004: 238–57) contrast OT with phonological systems that attempt to combine rules and constraints. Harmonic Grammar, another important predecessor of OT, has not been touched on here; see Prince and Smolensky (1993/2004: 234–8) for brief discussion and Legendre, Sorace, and Smolensky (2006) for a much more detailed treatment.

3. The Fundamentals of OT

How does OT respond to the issues raised in the previous section? That is, how does OT address the following questions?

(i) How are constraints on the output of the grammar satisfied? What is the relationship between constraints on output structures and the operations that transform inputs into outputs? How are triggering and blocking effects accounted for?
(ii) What is the relationship between the universal and the language particular? How can constraints differ in their activity from language to language?

The answers to these questions follow directly from the fundamental properties of OT.

OT sets up a basic dichotomy between the operational component of the grammar and the constraint component. The operational component, called GEN, constructs a set of candidate output forms that deviate from the input in various ways. The constraint component, called EVAL, filters the candidate set, selecting a member of it to be the actual output of the grammar.

GEN functions something like FM's optional transformations or GB's Move α. GEN applies all linguistic operations freely, optionally, and sometimes repeatedly. This property of GEN, which is known as **freedom of analysis**, is assumed for two reasons. First, it is simpler to define GEN with freedom of analysis than without it. For example, the phonological GEN can repeatedly epenthesize – that is, GEN(/pa/) includes {pa, pa**ə**, pa**əə**, pa**əəə**, . . . } – because complicating GEN with limits on epenthesis is unnecessary. Through minimal violation (see below) EVAL puts limits on epenthesis anyway. Second, GEN's freedom of analysis is necessary because of the related assumption that GEN is universal. Because GEN is the same in every language, it must in effect anticipate all of the ways that any language could transform a given input, so as to be certain that all of these options are represented in the candidate set. The simplest way to do this is to supply GEN with certain operations and allow them to apply freely, thereby over- rather than undergenerating the necessary range of candidates. Again, overgeneration by GEN does not mean overgeneration by the grammar, because the output of GEN is filtered by EVAL.

The grammar of a language is a constraint hierarchy. The constraint ranking relation is referred to as **domination**, written $\gg$. For example, the hierarchy [C1 >> C2], read as 'C1 dominates C2', is a statement about the relative priority of constraints C1 and C2 in a particular language.

EVAL receives the candidate set from GEN, evaluates it using the language's constraint hierarchy, and selects its **most harmonic** or **optimal** member as the output of the grammar. EVAL starts with the constraint that is ranked highest, C1, and it chooses only those candidates that are most favored by C1. The candidates that are most favored by C1 (or any other constraint) have two things in common: they receive the same number of violation marks from C1, and no other candidate receives fewer violation marks from C1. There is always at least one candidate that is most favored by C1, and it is possible for all of the candidates to be most favored by C1 (if all candidates violate C1 equally).

We now have a subset of the original candidate set consisting of just those candidates that are most favored by C1. This subset is passed along to the next constraint in the ranking, C2, and it does the same thing: it locates the subset of candidates that it most favors. This process continues until the set has been reduced to just one candidate. It is optimal. It satisfies the constraints as they are ranked in the language-particular hierarchy better than any other candidate in the original candidate set.

This brief explanation of EVAL illustrates several important points about OT:

(i) Although constraints are violable, violation is always **minimal**. The optimal candidate may violate C1, but no other candidate can violate C1 less than the optimal candidate does.
(ii) The preferences expressed by a lower-ranking constraint are important only insofar as they decide between candidates that tie on all higher-ranking constraints. In other words, OT constraints are arranged in **strict-domination hierarchies**, in which superior performance on lower-ranking constraints can never overcome inferior performance on higher-ranking constraints.
(iii) EVAL never asks for candidates that obey a constraint; it only asks for candidates that are favored by a constraint. This is an important distinction, because it means that there is always some optimal candidate (unless, absurdly, the initial candidate set is empty). Being favored by a constraint is not the same as obeying it. One or more candidates are always favored, but it will sometimes happen that no candidate obeys a given constraint.[2]

Each language has its own constraint ranking. The strongest hypothesis is that constraint ranking is the only thing in the grammar that is language particular: GEN, EVAL, and even the constraints themselves are universal. The universal constraint component is called CON. If the ranking of universal CON is the only difference between grammars, then all of the constraints in CON are present in the grammars of all languages. Only the ranking is different.

CON itself consists of two types of constraints. **Markedness** constraints are similar to the surface-structure constraints or filters of the 1970s. The inventory of markedness constraints in Con is a substantive theory of linguistic well-formedness – for example, complex consonant clusters or *that*-trace sequences are bad. A significant innovation in OT is the notion of a **faithfulness** constraint. Faithfulness constraints are inherently conservative, requiring the output of the grammar to resemble its input. Because markedness constraints favor some linguistic structures over others, they are often in tension with faithfulness constraints, which resist changes to input structures. This tension is called **constraint conflict**, and it is resolved in OT by ranking. Conflict between two markedness constraints or between two faithfulness constraints is also possible, of course.

These various aspects of OT are strongly interdependent. For example, faithfulness constraints make very little sense unless constraints are violable, because otherwise differences between input and output would be impossible.[3] Likewise, constraint violability and ranking are necessary for the claim that the constraints are universal, because otherwise languages could not differ.

This brief summary of the most important properties of OT is sufficient to support some preliminary answers to the questions raised at the outset of this section.

The first question concerns the interaction between linguistic processes and the output constraints that seem to trigger or block them. OT's answer is that there is no real interaction, as processes and constraints are in separate grammatical components, Gen and Eval, and information flows in only one direction, from Gen to Eval. Strictly speaking, processes are neither triggered nor blocked; instead, the process component Gen supplies a broad array of possible outputs that reflect the results of applying various operations. In Yawelmani, for example, unsyllabifiable consonants are resolved by epenthesis because candidates with epenthesis are among those supplied by Gen, and because Eval favors less-marked [ʔi.l**i**k.hin] over more-faithful *[ʔil.*k*.hin] and differently-unfaithful *[ʔil.hin]. Eval is responsible for *choosing* the winning candidate, not for *generating* it.

The second question asks how output constraints can differ in activity from language to language. OT attributes differences in constraint activity to differences in ranking. The ranking of a markedness constraint relative to faithfulness constraints and other markedness constraints determines whether and when it is active. We can illustrate this with some of the examples mentioned near the end of Section 2. Con includes a markedness constraint, called Onset, that requires syllables to have onsets. If Onset dominates the antiepenthesis faithfulness constraint (and certain other ranking requirements are met), then candidates that resolve onsetless syllables by epenthesis will be preferred to candidates that preserve them: Arabic /al-walad/ → [ʔal-walad], *[al-walad]. If in addition Onset is ranked below a faithfulness constraint specific to word-initial syllables, then epenthesis will occur medially but not initially, as in Axininca Campa: /i-N-koma-i/ → [iŋkoma**t**i] 'he will paddle'. On the other hand, if Onset is ranked below the antiepenthesis and antideletion faithfulness constraints, then we get a language like Timugon Murut that permits onsetless syllables initially and medially: /ambiluo/ → [am.bi.lu.o].

Even low rank is no guarantee that a constraint is completely inactive, however. Recall that Timugon Murut respects ONSET in two ways: intervocalic consonants are syllabified as onsets ([am.bi.**l**u.o], not *[am.bi**l**.u.o]); and the reduplicative prefix must have an onset ([a.**ba**.ba.lan], not *[**a**.a.ba.lan]). These effects of ONSET occur in situations where ONSET can be active without the need for epenthesis or deletion. Situations like this, where even a low-ranking markedness constraint can be decisive in situations where faithfulness is not at issue, are known as **emergence of the unmarked** (McCarthy and Prince 1994).

As these few examples make clear, constraint ranking gives much more subtle control over constraint activity than parameters do. These examples also remind us, as we saw in Section 2, that more subtle control is exactly what is needed.

Suggestions for further reading. The fundamentals of OT are laid out primarily in Chapter 5 of Prince and Smolensky (1993/2004). Other overviews of this topic, in approximate order of increasing length and depth of coverage, include Archangeli (1997), Tesar, Grimshaw, and Prince (1999), Kager (1999: 1–51), Prince and Smolensky (1997; 2003), McCarthy (2002: 3–47; 2003a), and Smolensky, Legendre, and Tesar (2006). Legendre (2001) is an introduction to OT syntax.

4. Doing Linguistic Analysis in Optimality Theory

As constraint ranking is hypothesized to be the only way in which grammars may differ, the core of any OT analysis is a collection of constraint rankings and the justification for them. In general, a ranking is justified by comparing two output candidates from the same input. One of these candidates, called the winner, is the actually observed output for that input in the language under analysis. The other candidate is a loser: it is derived by GEN from that same input, but it is not the most harmonic candidate according to EVAL. For EVAL to select the correct candidate as winner, *some* constraint that favors the winner over the loser must dominate *every* constraint that favors the loser over the winner. The logic of this statement follows from the properties of EVAL (see Section 3). Constraint **ranking arguments** depend on this logic.

There are three essential elements to a valid ranking argument: First, the constraints to be ranked must conflict; that is, they must disagree in their assessment of a pair of competing output candidates derived from the same input. For instance, Yawelmani /ʔilk-hin/ has competing candidate output forms [ʔi.l**i**k.hin] and *[ʔil.*k*.hin]. The markedness constraint against unsyllabified consonants (call it **C*, recalling the italicization of unsyllabified consonants) and the antiepenthesis faithfulness constraint (known as DEP) disagree in their assessment of these forms.

Second, one member of this pair of competing candidates must be the actual output form for the given input – the winner, in short. The constraint that favors the winner, *C, must dominate the constraint that favors the loser, DEP. Third, the ranking argument is secure only if there is no third constraint that could also be responsible for this winner beating this loser. In the Yawelmani example, such a constraint would have to meet two conditions to be problematic: like *C, it would have to favor [ʔi.l**i**k.hin] over *[ʔil.*k*.hin]; and it could not be ranked below *C. Suppose for the sake of argument that there were a constraint *lk that is violated by any [lk] consonant sequence. Because *lk prefers [ʔi.l**i**k.hin] over *[ʔil.*k*.hin], just as *C does, *lk threatens to undermine the ranking argument we have constructed, leaving us with a less definitive disjunction: at least one of *C and *lk dominates DEP. This disjunction can be resolved by independently showing that DEP dominates *lk, using a ranking argument based on the winner/loser pair [ʔil.kal]/*[ʔi.l**i**.kal] 'might sing'.

The elements of a ranking argument are illustrated with a **tableau**. The original **violation tableau** format of Prince and Smolensky (1993/2004) is illustrated in (1). Each cell (*row*, *column*) indicates with asterisks the number of violations, if any, of constraint *column* incurred by candidate *row*. When a constraint knocks a candidate out of competition, that result is signaled by an exclamation point. Cells are shaded when they can have no possible effect on the outcome because higher-ranking constraints are decisive.

(1) Violation tableau

/ʔilk-hin/	*C	DEP
a.→ ʔi.l**i**k.hin		*
b. ʔil.*k*.hin	*!	

The **comparative tableau**, introduced by Prince (2002), adds annotations to each loser row indicating whether the constraint favors that loser or the winner. In the comparative tableau (2), *C favors the winner over the loser [ʔil.*k*.hin], so a W appears in the *C cell of the loser row. DEP favors the loser, so an L is placed in the DEP cell. In a properly ranked comparative tableau, every loser row contains at least one W with no Ls to its left (readers may find it helpful to pause at this point and convince themselves that the previous statement follows from the nature of EVAL).

(2) Comparative tableau

/ʔilk-hin/	*C	DEP
a. → ʔi.l**i**k.hin		*
b. ʔil.*k*.hin	*W	L

The comparative tableau is best for dealing with the **ranking problem**, as Brasoveanu and Prince (2005) call it. In the ranking problem, the winner is already known and we are trying to figure out which ranking will produce that winner. The ranking problem arises whenever we are trying to analyze some data: the winners are known because they are the data of the language, and it is the ranking that we seek to discover. The violation tableau format is appropriate for what Brasoveanu and Prince call the **selection problem**. In the selection problem, the ranking is already known but the winner is not. This is a typical situation in the study of language typology, where we want to check each ranking to determine which candidate wins.

Every theory presents its own special challenges to the analyst, and OT is no exception. One rather difficult task is determining the range of candidates that must be considered. Overlooking a losing candidate that ties or beats the intended winner 'invites theoretical disaster, public embarrassment, and unintended enrichment of other people's careers' (McCarthy and Prince 1993: 13), as the overlooked candidate has the potential to undermine the entire analytic edifice. Systematic exploration of some space of candidates is usually the best way to avoid this problem.[4] For example, once we have established that DEP must be outranked by other constraints in Yawelmani, it makes sense to consider all of the ways in which GEN can alter /ʔilk-hin/ by epenthesis. We can quickly determine that candidates with absurdly repeated epenthesis like *[ʔi.l**i**k.hi.n**i**.**i**.**i**.**i**] are no threat, as they violate DEP more than the winner does and do not beat it on any known constraint. But a candidate like *[ʔi.l**i**.k**i**.hi.n**i**] has a more serious claim on our attention. It discloses that DEP must dominate the markedness constraint NO-CODA, which prohibits syllable-final consonants. Even more important is the candidate *[ʔil.k**i**.hin], because it ties with the intended winner [ʔi.l**i**k.hin] on both **C* and DEP. The analysis would not be complete without some constraint that decided between these two candidates.

Another special responsibility of the analyst in OT is the introduction of new constraints. In theory, CON is universal, so the analyst's job is just a matter of finding some ranking of CON that reproduces the data of the language in question. In practice, however, the details of the constraints in CON are the topic of ongoing research and discussion. Because CON embodies all of the substantive properties of human language, our present inability to deliver a definitive theory of CON should come as no surprise to anyone willing to concede that there are things about language that are not yet fully undersood.

What this means for the analyst is that any actual analysis may very well involve some constraints whose status in universal grammar (i.e., CON) is not yet settled. For example, while there is probably a consensus among phonologists that CON includes a faithfulness constraint MAX that militates against deleting elements of the input, there is disagreement about whether the antiepenthesis constraint Dep is also needed, as many of its effects can be obtained from independently necessary markedness constraints [see Gouskova (2007) and Urbanczyk (2006) for opposing views on this question]. Another example: different ways of rationalizing the ad hoc markedness constraint *C depend on different theories of the relationship between syllabification and phonotactics. OT itself says nothing about syllabification or phonotactics, although it certainly provides a framework in which a theory of syllabification and phonotactics can be worked out using violable constraints.

OT provides the analyst with ample opportunities to test and refine any hypothesis about Con. Any newly proposed constraint must, of course, offer insight into the analysis at hand. But this is not enough; the proposer is also responsible for exploring the (un)intended consequences of the new constraint under ranking permutation. The source of this novel obligation on the analyst is OT's inherently typological character: every language-particular hierarchy of n constraints is pregnant with $n!-1$ other hierarchies, each of which is predicted to be the grammar of a possible human language.[5] Fortunately, not all constraints conflict, so many different rankings will produce identical results, making the task of the analyst-*cum*-theorizer far less daunting than it might at first seem.

As an example of this sort of reasoning, we can return to a question raised just above: Is the faithfulness constraint Dep unnecessary because independently motivated markedness constraints do the same work? Take, for instance, a language like Timugon Murut that allows onsetless syllables. In this language, Onset must be crucially dominated to rule out consonant epenthesis. Either Onset is dominated by Dep or, in the Dep-less theory, Onset is dominated by a set of markedness constraints that is sufficient to rule out every imaginable consonant that could be inserted to satisfy Onset. For convenience, we can refer to this set of constraints as *Cons. The Dep-less theory requires the existence of *Cons, and therefore it makes certain predictions about the effects of ranking permutation. For example, by ranking *Cons above Max, we could get a language that has no consonants whatsoever, only vowels (Gouskova 2003: 71ff). No such language is known, and nearly everyone would agree that we do not want a theory to predict the existence of such a language. In this case as in so many others, tests against language typology act as a check and corrective on proposals about Con (see Section 6.1 for further discussion.)

Suggestions for further reading. The special challenges of analysis in OT are briefly addressed in McCarthy (2002: 30–42). McCarthy (forthcoming) is a book-length treatment of this topic.

5. Illustrations

This section puts some flesh on the theoretical and analytic skeleton assembled in the previous sections. It does this by illustrating OT at work with two examples, one phonological and the other syntactic.

In the Nigerian Niger-Congo language Emai (Casali 1996: 62–8; Schaefer 1987), ONSET is satisfied at V_1+V_2 juncture by changing V_1 into a glide, deleting V_1, or deleting V_2. The choice among these three options depends on whether V_1 is a high vowel and on whether the morphemes that contain V_1 and V_2 are lexical or functional:[6]

(3) Synopsis of Emai alternations
At V_1+V_2 juncture:

a. If V_1 and V_2 are both contained in lexical morphemes and V_1 is a high vowel [i] or [u], then V_1 changes into the homorganic glide [j] or [w]:

/k**u** amɛ/	[k**w**amɛ]	'throw$_{Lex}$ water$_{Lex}$'
/f**i** ɔpia/	[f**j**ɔpia]	'throw$_{Lex}$ cutlass$_{Lex}$'

Otherwise:

b. If one of the morphemes is lexical and the other is functional, delete the vowel in the functional morpheme:

/ɔl**i** ebe/	[ɔlebe]	'the$_{Func}$ book$_{Lex}$'
/ukpode **ɔ**na/	[ukpodena]	'road$_{Lex}$ this$_{Func}$'

c. If both morphemes are lexical or both are functional, delete V_1:

/k**ɔ** ema/	[kema]	'plant$_{Lex}$ yam$_{Lex}$'
/f**a** edi/	[fedi]	'pluck$_{Lex}$ palm-nut$_{Lex}$'
/oa is**i** ɔi/	[oasɔi]	'house$_{Lex}$ of$_{Func}$ his$_{Func}$'

The faithful candidates derived from these inputs violate ONSET: *[ku.a.mɛ], etc. The main analytic challenge is determining how ONSET is satisfied, and that is primarily a matter of determining the correct ranking of certain faithfulness constraints. One of them might be called IDENT(syllabic); it is violated when a vowel is replaced by a glide. The others are versions of the antideletion constraint MAX. In Emai, as in various other languages, the roots of major lexical category items (nouns, verbs, adjectives) are treated more faithfully than function words or affixes. This evidence shows that CON must contain a constraint MAX$_{Lex}$ that is violated only by deletion of segments from noun, verb, or adjective roots (McCarthy and Prince 1995). MAX$_{Lex}$ is distinct and separately rankable from the general constraint MAX, which is violated by deletion of any segment, regardless of the lexical status of the morpheme that contains it. Similarly, many languages show greater faithfulness to segments in initial syllables, motivating the positional faithfulness constraint MAX$_{Init}$ (Beckman 1997, 1999; Casali 1996). It too is distinct and separately rankable from other MAX constraints.

With these typologically justified constraints in hand, we are now in a position to establish the ranking responsible for the alternations in (3). We will begin with the process of glide formation. Changing a vowel to a glide is a violation of the faithfulness constraint IDENT(syllabic). Whenever an output of the grammar violates a faithfulness constraint, some markedness constraint must dominate that faithfulness constraint. In this case, the high-ranking markedness constraint is ONSET:

(4) ONSET >> IDENT(syllabic)

/ku$_{Lex}$ amɛ$_{Lex}$/	ONSET	IDENT
a.→ kwa.mɛ		*
c. ku.a.mɛ	*W	L

Deleting a vowel is an obvious alternative to changing it into a glide, and that possibility must be excluded with another constraint, also ranked above IDENT(syllabic). For reasons that will become clear later, that role belongs to MAX$_{Lex}$ and not to either of the related constraints MAX$_{Init}$ and plain MAX:

(5) MAX$_{Lex}$, ONSET >> IDENT(syllabic)

/ku$_{Lex}$ amɛ$_{Lex}$/	ONSET	MAX$_{Lex}$	IDENT
a.→ kwa.mɛ			*
b. ka.mɛ		*W	L
c. ku.a.mɛ	*W		L

Given just the ranking in (5), we might expect to see other situations where glide formation is preferred to deletion, such as the following:

(6) e]$_{Lex}$ [$_{Lex}$a → ja
a]$_{Lex}$ [$_{Lex}$i → aj

Neither of these alternations occurs, so some constraints must rule them out. The mapping /e a/ → [ja] violates an undominated faithfulness constraint, IDENT(high). The output [aj] violates an undominated markedness constraint against diphthongs with falling sonority, such as [aj] and [aw].

Although a high vowel at the end of a lexical morpheme can become a glide because of the ranking in (5), a high vowel at the end of a functional morpheme deletes instead: /ɔli$_{Func}$ ebe$_{Lex}$/ → [ɔ.le.be], *[ɔ.lje.be]. Deleting the /i/ of /ɔli/ is a violation of general MAX, but not of MAX$_{Lex}$. Therefore, as (7) shows, IDENT(syllabic) dominates MAX. With this ranking, deletion is preferable to a vowel→glide mapping, as long as the contents of lexical morphemes are not threatened (on the winner's remaining ONSET violation, see Note 6).

(7) IDENT(syllabic) >> MAX

/ɔli$_{Func}$ ebe$_{Lex}$/	ONSET	MAX$_{Lex}$	IDENT	MAX
a. → ɔ.le.be	*			*
b. ɔ.lje.be	*		*W	L
c. ɔ.li.e.be	**W			L

From the ranking arguments in (4), (5), and (7), we can assemble the constraint hierarchy [MAX$_{Lex}$, ONSET >> IDENT(syllabic) >> MAX]. According to this hierarchy, ONSET is satisfied by deleting the vowel of a functional morpheme, or else changing the vowel of a lexical morpheme into a glide. In this hierarchy, MAX$_{Lex}$ and ONSET are not yet ranked with respect to one another. Ranking requires conflict, and the conditions that will place MAX$_{Lex}$ and ONSET into conflict are quite precise: both vowels must be in lexical morphemes, and glide formation must be ruled out for the reasons given just below (6). An example that meets both of these criteria is /kɔ$_{Lex}$ ema$_{Lex}$/ → [ke.ma]. It shows that the vowels of lexical morphemes can delete, so ONSET must dominate MAX$_{Lex}$:

(8) ONSET >> MAX$_{Lex}$

/kɔ$_{Lex}$ ema$_{Lex}$/	ONSET	MAX$_{Lex}$
a. → ke.ma		*
b. kɔ.e.ma	*W	L

The /kɔ$_{Lex}$ ema$_{Lex}$/ → [ke.ma] mapping in (8) and other examples raise another question: when glide formation is not an option, which vowel is deleted? The facts are as follows. If one vowel is hosted by a lexical morpheme and the other is hosted by a functional morpheme, then the vowel of the lexical morpheme is preserved because of MAX$_{Lex}$: /ukpode$_{Lex}$ ɔna$_{Func}$/ → [ukpodena]; /ɔli$_{Func}$ ebe$_{Lex}$/ → [ɔ.le.be]. If both vowels are hosted by lexical morphemes or both are hosted by functional morphemes, then MAX$_{Init}$ favors preservation of the vowel that is morpheme initial: /kɔ$_{Lex}$ ema$_{Lex}$/ → [ke.ma].

Interestingly, although MAX$_{Init}$ is crucial for the analysis of /kɔ$_{Lex}$ ema$_{Lex}$/ → [ke.ma], this example does not tell us how MAX$_{Init}$ is ranked. As tableau (9) shows, MAX$_{Init}$ has the role of a tie-breaker: no other constraint favors one of the candidates. This means that MAX$_{Init}$ does not conflict with any of the other constraints over this pair of candidates [Observe that there is no L annotation in row (b)]. As conflict is the basis of ranking arguments, tableau (9) offers no evidence about how MAX$_{Init}$ is ranked.

(9) MAX$_{Init}$ as tie-breaker

/kɔ$_{Lex}$ ema$_{Lex}$/	ONSET	MAX$_{Lex}$	IDENT	MAX		MAX$_{Init}$
a. → ke.ma		*		*		
b. kɔ.ma		*		*		*W

MAX$_{Init}$ is rankable on the basis of other data, however. To get a conflict and therefore a ranking argument, we need an example where the choice is between deleting a Lex vowel that is noninitial and deleting a Func vowel that is initial. Examples like /ukpode$_{\text{Lex}}$ ɔna$_{\text{Func}}$/ → [ukpodena] are just what is required. As the Func-initial vowel is deleted and the Lex-final vowel is not, Max$_{\text{Lex}}$ must dominate Max$_{\text{Init}}$:

(10) Max$_{\text{Lex}}$ ≫ Max$_{\text{Init}}$

/ukpode$_{\text{Lex}}$ ɔna$_{\text{Func}}$/	Max$_{\text{Lex}}$	Max$_{\text{Init}}$
a. → u.kpo.de.na		*
b. u.kpo.dɔ.na	*W	L

The diagram in (11) illustrates the rankings that have been established. In such a diagram, if there is a strictly downward path from A to B, then A dominates B by direct argument or by transitivity of the domination relation.

(11)

$$\begin{tikzcd}[column sep=small] & \textsc{Onset} \arrow[d, no head] & \\ & \textsc{Max}_{\text{Lex}} \arrow[dl, no head] \arrow[dr, no head] & \\ \textsc{Max}_{\text{Init}} & & \textsc{Ident}(\text{syllabic}) \arrow[d, no head] \\ & & \textsc{Max} \end{tikzcd}$$

In the parlance of 1970s phonological theory, Emai would be said to have a **conspiracy** (a term originally due to Haj Ross). In a conspiracy, several distinct rules act in concert to achieve the same output result (Kisseberth 1970). Emai would require at least three rules – glide formation, deletion of the first vowel, and deletion of the second vowel – all in service of eliminating onsetless syllables. A conspiracy is really just a description of something that rule-based phonology is unable to explain. In OT, on the other hand, there are no rules and hence no conspiracy. The output constraint Onset favors candidates without onsetless syllables, and it dominates all four faithfulness constraints in (11). Ranking relations among the various faithfulness constraints determine which unfaithful candidates are the winners. The output constraint Onset is obviously central to the analysis and not just a post hoc rationalization of why this particular language should have these three particular processes.

Grimshaw's (1997) account of *do*-support in English illustrates a very different kind of analysis in OT. The goal is to explain the economy of unstressed *dŏ*: why does it appear only when it is needed? Unstressed *dŏ* is required in inversion and negation constructions but forbidden in simple declaratives:

(12) a. Dĭd Robin leave?
b. When dĭd Robin leave?
c. Robin dĭdn't leave.
d. *Robin dĭd leave.

Dŏ's only-when-needed distribution means that there must be a markedness constraint that militates against it, to force its absence from surface structure except when some higher ranking constraint compels its presence. For Grimshaw, the constraint with anti-*dŏ* effect is FULL INTERPRETATION (FULL-INT), which is violated by any lexical item in surface structure that does not contribute to the interpretation of that structure.

FULL INTERPRETATION is sometimes in conflict with the constraint OBLIGATORY HEADS (OB-HD), which is violated by any projection that lacks a head. When a projection would be headless, the choice is between supplying a head that violates FULL INTERPRETATION or tolerating the breach of OBLIGATORY HEADS. In English, *do*-support is the result of a constraint hierarchy in which OBLIGATORY HEADS ranks higher:

(13) OBLIGATORY HEADS ≫ FULL INTERPRETATION

	OB-HD	FULL-INT
a. → [CP When dĭd$_i$ [IP Robin e$_i$ [VP leave t]]]		*
b. [CP When e [IP Robin e [VP leave t]]]	**W	L
c. [CP When e [IP Robin dĭd [VP leave t]]]	*W	*

Although it is ranked below OBLIGATORY HEADS and does not affect the outcome in (13), FULL INTERPRETATION is not inactive. There are situations, such as simple declaratives, where every projection has a head without further ado, so to speak. In that case, FULL INTERPRETATION emerges to ban *dŏ* from the output:

(14) Economy of *dŏ*

	OB-HD	FULL-INT
a. → [VP Robin left]		
b. [IP Robin dĭd [VP leave]		*W

For the same reason, *dŏ* is impossible in combination with another auxiliary: **When will Robin dŏ leave?*

Exactly parallel situations are found in phonology (Gouskova 2003): in Lillooet, for instance, the vowel [ə] appears only when it is needed for markedness reasons, in words that would otherwise be vowelless and in clusters that would otherwise violate sonority-sequencing requirements. Grimshaw and Gouskova show, for syntax and phonology respectively, how such economy effects follow from intrinsic properties of OT rather than stipulated economy principles. With English *do*-support, and likewise with Lillooet [ə], the source of the economy effect is minimal violation of constraints like FULL-INT. Minimal constraint violation is a *sine qua non* of OT; it is inherent to OT's core component, EVAL (see also Section 6.2).

6. Some Questions and Answers about OT

Many of the most commonly asked questions about OT touch on fundamental aspects of the theory. These questions therefore serve as an excellent framework for further exploration of OT's concepts and consequences [for other lists of frequently asked questions about OT, see McCarthy (2002: 239–45), Smolensky, Legendre, and Tesar (2006: 523–32), and Prince and Smolensky (1993/2004: 232)].

6.1 LANGUAGE UNIVERSALS

Because all constraints are violable, how does OT account for universals of human language? In general, what sorts of predictions does OT make about possible and impossible languages?

In answering this question, it is helpful to begin with a distinction due to Prince (1997a). Some predictions follow from just OT's basic properties, while other predictions follow from OT's basic properties under some specific assumptions about the contents of Con. Universals of the first type can be called 'formal', and those of the latter type can be referred to as 'substantive'.

The simplest example of a formal universal in OT is **harmonic improvement**. As Con contains only markedness and faithfulness constraints, a necessary condition for an unfaithful candidate to win is that it be less marked than the faithful candidate – in other words, an unfaithful winner must improve harmonically over a faithful loser. This property follows from the basic structure of OT. If the input is /bi/, then no faithfulness constraint disfavors faithful [bi] as the output. For unfaithful [be] to win, then, some higher-ranking markedness constraint M must favor [be] over [bi]. OT's basic assumptions entail that unfaithfulness is possible only when it improves markedness, where improvement is measured relative to the universal markedness constraints as they are ranked in the language in question.

Moreton (2003) develops a formal proof of this result and explores its empirical consequences. One is that no OT grammar can analyze a language with a system of circular mappings like /bi/ → [be] and /be/ → [bi]. The reason is that no single constraint ranking can simultaneously assert that [be] is less marked than [bi] (to get the /bi/ → [be] mapping) and [bi] is less marked than [be] (to get the /be/ → [bi] mapping).[7] Another consequence is that no OT grammar can analyze a system of unbounded growth, like /bi/ → [biʔi], /biʔi/ → [biʔiʔi], . . . The reason is that a point will always be reached where further addition of syllables is not harmonically improving.

Substantive universals are mainly concerned with language typology. Everything in OT is invariant across languages – Gen, Eval, Con, and even the inputs (see Section 6.3) – except for the constraint ranking.

Therefore, the various ways of ranking Con define the permitted range of between-language variation. This is **factorial typology**: the claim that every permutation of Con is a possible grammar of a human language, and every actual human language has a grammar that is a permutation of Con. Claims about typology are substantive universals because they depend in an obvious way on what constraints Con does and does not contain.

Consider a simple substantive universal like the fact that no language satisfies Onset by epenthesizing a consonant cluster, mapping, for example, /apati/ to [trapati]. If [trapati] is in /apati/'s candidate set, as is standardly assumed (McCarthy and Prince 1995, 1999), then the explanation for this universal requires that [trapati] not be the optimal member of /apati/'s candidate set under any permutation of Con. This mapping is correctly predicted to be universally impossible if, say, Con does not include a markedness constraint that favors initial consonant clusters over single initial consonants. This same claim about Con makes other typological predictions as well. For instance, there can be no language that requires all of its syllables to begin with a cluster, nor can there be a language that limits reduplication to cluster-initial words.

Because substantive universals depend on specific assumptions about Con, disagreement among analysts is expected. In fact, typology is an essential check on all hypothesized changes to the make-up of Con. No newly proposed constraint is secure until its consequences under ranking permutation have been checked for typological plausibility (see Section 4).

Suggestions for further reading. For further discussion of formal universals in OT, see Prince (1997a) and McCarthy (2002: 109–11). Almost any of the suggested readings at the end of Section 3 is a good place to learn more about factorial typology and substantive universals. The use of factorial typology to draw or confirm inferences about Con is (or at least should be) ubiquitous in the OT literature.

6.2 constraints versus parameters

How is constraint ranking different from constraint parameterization? Isn't a low-ranking constraint effectively turned off, just like a parameter?

This issue was already touched on at the end of Section 2, but it is important enough to bear repeating. Crucial domination of a constraint is no guarantee of inactivity. There are some very special circumstances where inactivity can be guaranteed (Prince 1997b: 3; Prince and Smolensky 1993/2004: 97–9, 130–2, 264–5), but otherwise a low-ranking constraint is always *potentially* active. A low-ranking constraint will be active *in fact* when it decides among candidates that tie on all higher-ranking constraints and perform better than all other candidates on those constraints. This is decidedly unlike the behavior of parameterized constraints: when such a constraint is turned off, it is never active under any circumstances.

Constraints that are low-ranking but active are particularly important in OT's explanation for economy effects. Faithfulness constraints ensure a kind of economy of derivation. Even when a faithfulness constraint is crucially dominated, such as Dep in Yawelmani, violation is always minimal, and so gratuitous epenthesis is ruled out: /ʔilk-hin/ → [ʔi.**li**k.hin], *[ʔi.l**i**.k**i**.hi.n**i**]. Certain markedness constraints effect a kind of economy of representation in both syntax and phonology (see Section 5).

There are other, more subtle differences between ranking and parametrization. To cite a phonological example that also has syntactic analogs, consider structures of the form $[\ldots(\ldots)_{\text{foot}}\ldots]_{\text{word}}$. There are constraints that require every foot to be aligned at the left or right edge of the word, Align-L(foot, word) and Align-R(foot, word). Align-L(foot, word) favors [(ˈpata)kama] over [pa(ˈtaka)ma] and [pata(ˈkama)]. Align-R(foot, word) favors [pata(ˈkama)] over [(ˈpata)kama] and [pa(ˈtaka)ma]. It might seem, then, that [(ˈpata)kama] or [pata(ˈkama)] are the only possible winners, depending on whether Align-L(foot, word) or Align-R(foot, word) is ranked higher. But this quasi-parametric conception of ranking overlooks candidates like [(ˈpata)], which manage to satisfy both of the alignment constraints at the expense of eliminating some input material. For a real-life example, reduplication in Diyari, see McCarthy and Prince (1994).

Cases like this are particularly relevant to the comparison of OT with parametric theories, as Samek-Lodovici (1996: 216) argues. In a parametric theory of stress like Hayes (1995), the direction of foot assignment is a parameter with two values, left-to-right and right-to-left. Nothing further follows from this. In OT, however, the constraints Align-L(foot, word) and Align-R(foot, word) together entail a further possibility: satisfaction of both constraints in words consisting of exactly one foot. Here, as in many other cases, OT's nonparametric approach to between-language variation has economy of structure as an unavoidable – and welcome – side effect.

Suggestions for further reading. The difference between ranking and parameterization is a particular focus of Grimshaw (2002), McCarthy (2002: 11–2, 109–10, 127–08 et passim), McCarthy and Prince (1994), Prince and Smolensky (1993/2004: 164–7), and Samek-Lodovici (1996), among others.

6.3 THE LEXICON

What is the role of the lexicon in OT?

To a great extent in syntax, and to a lesser extent in phonology, the lexicon has been seen as the proper repository of all that is language particular. OT turns that assumption on its head.

One of OT's basic premises is called **richness of the base** (Prince and Smolensky 1993/2004: 205, 225). The 'base' is the set of inputs to the grammar [an allusion to the base component of Chomsky (1957)]. It is 'rich' because, by hypothesis, it is not subject to any language-particular restrictions. In phonology, richness of the base rules out the use of morpheme structure constraints, stipulated underspecification, or similar devices that preemptively remove forms from the set of possible inputs to the grammar. In syntax, richness of the base means that systematic differences between languages cannot be attributed to systematic differences in the contents of their lexicons. Because of richness of the base, all aspects of well-formedness are under the control of EVAL and the constraint hierarchy, and all systematic differences between languages must be obtained from differences in constraint ranking. Richness of the base is also central to OT's resolution of the so-called duplication problem in phonology (McCarthy, 2002: 71–6).

Richness of the base has been a source of some confusion in criticisms of OT. It does not mean, absurdly, that all languages have the same vocabulary, nor does it mean, equally absurdly, that learners posit something like /ʕɠ̊kæt/ the underlying form for [kæt] *cat*. Rather, it simply means that the lexicon as a system is not subject to any language-particular requirements. Furthermore, it means that explanations for linguistic phenomena cannot involve carefully contrived limits on the inputs to the grammar. An example from phonology: between-language differences in the behavior of laryngeal features must derive from differences in ranking rather than differences in lexical specification of [voice] and [spread glottis] (Beckman and Ringen 2004). An example from syntax: between-language differences in fronting of *wh* words must derive from differences in ranking rather than differences in whether *wh* words are lexically specified with a strong or weak feature (Smolensky, Legendre and Tesar 2006: 529).

Suggestions for further reading. As richness of the base is ubiquitous in the OT phonology literature, I focus here on syntax. Richness of the base in syntax receives special attention from Aissen (1999), Bakovic and Keer (2001), Bresnan (2001a,b), Grimshaw (1997, 2001), Grimshaw and Samek-Lodovici (1995, 1998), Legendre, Smolensky, and Wilson (1998), and Samek-Lodovici (1996: 219).

6.4 UNGRAMMATICALITY

In a theory where all constraints are violable, how can any linguistic representation be absolutely ill formed? For example, how is it possible for English phonology to rule out *[bnɪk] as phonotactically impossible, and how is it possible for English syntax to rule out **Who did he say that left?* as ungrammatical?

Although other theories use inviolable constraints and crashing derivations to account for ungrammaticality, these analytic techniques are not indispensable. In OT, ungrammaticality is a consequence of inferiority to other candidates rather than an apparatus of inviolable constraints. The linguistic representation *[A] is ungrammatical in some language if and only if [A] is not the optimal candidate for any input, given that language's constraint hierarchy. That is, for any input /X/ that has [A] in its candidate set, there is some output candidate that is more harmonic than [A] according to the grammar of the language in question.

An input that merits particular attention is /A/ itself, such as /bnɪk/. As every faithfulness constraint favors the mapping /bnɪk/ → [bnɪk], some higher-ranking markedness constraint is necessary to rule it out in English. This constraint is presumably a prohibition on onset clusters of two stops, and if it dominates Dep, then [bənɪk] will be more harmonic than *[bnɪk] ([bənɪk] is not a real word of English, but it is phonotactically possible, and that is the point of the example). This is not quite enough to secure the absolute ill-formedness of *[bnɪk], however; that requires showing that *no* input will map most harmonically to *[bnɪk]. That demonstration ultimately requires using methods of proof rather than exhaustive search, as there are infinitely many possible inputs.

Suggestions for further reading. Absolute ill-formedness is the topic of an anthology, Rice (forthcoming). Earlier work that focuses specifically on this topic includes Bakovic and Keer (2001), Legendre, Smolensky, and Wilson (1998: 257), McCarthy (2002: 76, 176, 194–9, 239), Orgun and Sprouse (1999), Pesetsky (1997), Prince and Smolensky (1993/2004: 57ff), Raffelsiefen (2004), Rice (2003, 2005), and Smolensky, Legendre, and Tesar (2006: 528–30).

6.5 INFINITY

If the number of candidates is infinite or even very large, then how do speakers find the optimal candidate in a reasonable length of time? For instance, the phonological Gen described in Section 3 can iterate epenthesis forever, so the candidate set is unbounded in size.

This question is based on two tacit assumptions. One of them is that some other theory of language is more plausible psychologically than OT. Presumably, we are to imagine a speaker mentally applying a sequence of rules before producing an utterance and to compare this with the image of a speaker hopelessly trying to evaluate an infinite set of candidate utterances. The problem is that the image of the speaker applying a sequence of rules has just as little claim to psychological plausibility. The failure of the Derivational Theory of Complexity shows this, at least in syntax (Fodor, Bever and Garrett 1974); the same seems to be true in phonology as well (Goldsmith 1993).

The other assumption underlying this question is an implicit denial of the competence/performance distinction. In other words, the question presupposes that a generative grammar is a model of a speaker's actual processing behavior as well as a speaker's linguistic knowledge. Yet this is not the stated goal of generative grammar. For example, Chomsky (1968: 117) writes that '. . . although we may describe the grammar *G* as a system of processes and rules that apply in a certain order to relate sound and meaning, we are not entitled to take this as a description of the successive acts of a performance model . . . – in fact, it would be quite absurd to do so. . . . If these simple distinctions are overlooked, great confusion must result.'

Issues of efficient generation and parsing are of course important, but they do not bear directly on the theory of competence. They are properly addressed using the ideas and methods of fields like computational linguistics. See the suggested readings for some examples.

Suggestions for further reading. There has been a great deal of research and progress on computational modeling of OT and determining its computational complexity. Tesar (1995) is probably the earliest work on these topics, and there is much more (Ellison 1994; Eisner 1997; Hammond 1997; Frank and Satta 1998; Gibson and Broihier 1998; Karttunen 1998; Wareham 1998; Fanselow et al. 1999; Gerdemann and van Noord 2000; Riggle 2004; Becker 2006; and others).

6.6 LEARNING

The number of permutations of n constraints is $n!$. How then is it possible for learners to find the right ranking for their language out of the gigantic space of possible rankings?

The ranking is actually the easiest part of language learning in OT. Given the right kind of data, the constraint demotion learning algorithm rapidly and efficiently finds a ranking that will correctly reproduce the data (Tesar and Smolensky 1998, 2000). The basic idea is this: every constraint that favors a loser over the intended winner must be dominated by some constraint that favors the winner over that loser (this is easy to visualize in a comparative tableau: every L must have some W to its left). This requirement follows from the nature of EVAL, and it is the basis of ranking arguments like those in Section 4. Learning proceeds by demoting every loser-favoring constraint below some winner-favoring constraint, until all loser-favoring constraints are crucially dominated. The losers that drive this learning algorithm can be obtained from the grammar itself: they are outputs produced by the incorrectly ranked grammar before learning is complete.

The data needed for the constraint demotion learning algorithm is more than just the phonetic signal that the learner hears; it must also contain information about hidden structure. In both syntax and phonology, learners may have multiple options for parsing expressions of the ambient language, and getting the parse right is crucial to getting the ranking right. In phonology, the ranking also depends on the accurate recovery of underlying representations from paradigmatic alternations. There has been significant progress on the hidden structure problem in OT; the main idea is that learners use their nascent grammar to produce hypotheses about hidden structure, and further learning reveals whether these hypotheses are correct (Tesar 1998, 1999; Tesar and Prince 2004).

A particularly welcome consequence of OT is its success in making connections among phonological theory, learnability, and empirical research on language acquisition. With the exception of Natural Phonology (Stampe 1973; Donegan and Stampe 1979), pre-OT generative phonology was confounded by the facts of language acquisition: children's reduced pronunciations required that child phonology have many rules for which there is no evidence in the adult language. In OT, children's reduction processes are a result of satisfying high-ranking universal markedness constraints. The very same markedness constraints that, through ranking, characterize differences between languages are also responsible for differences between children and adults within a single language.

Suggestions for further reading. There are two anthologies that are relevant to this topic: Kager, Pater, and Zonneveld (2004) and Dinnsen and Gierut (forthcoming). The former includes work on learnability as well as acquisition. The latter deals with disordered as well as normal acquisition of phonology. McCarthy (2002: 232) has a nearly exhaustive list, compiled by Joe Pater, of the OT literature on acquisition before 2002.

6.7 DERIVATIONS

In OT, inputs are mapped to outputs without any intermediate steps. Aren't there linguistic phenomena, in both phonology and syntax, that require derivations with intermediate steps (Chomsky 1995: 223–5, 380 and many others)?

It is quite correct that the mapping from the input to the output of an OT grammar does not involve any intermediate steps. For example, the mapping from underlying /ktub/ to surface [ʔuktub] 'write!' in Arabic involves a two-step derivation in rule-based phonology (15), with vowel epenthesis creating the context that necessitates [ʔ] epenthesis. In OT, on the other hand, the grammar compares candidates that may show the simultaneous effects of two or more epenthesis operations, and [ʔuktub] is among them (16).

(15) Arabic /ktub/ → [ʔuktub] with rules

Underlying	/ktub/
Vowel epenthesis	uktub
[ʔ] epenthesis	ʔuktub
Surface	[ʔuktub]

(16) Arabic /ktub/ → [ʔuktub] in OT

/ktub/	Onset	*#CC	Dep
a.→ ʔuktub			**
b. ktub		*W	L
c. uktub	*W		*L

Why does OT have flat derivations? There are empirical arguments (many of which are summarized in McCarthy 2002: 138–63), but the main reason is theoretical parsimony. Nonflat derivations are often a way of establishing priority relationships among linguistic requirements, and OT already has a way of setting priorities, ranking. For example, in the Wakashan language Nuuchahnulth (formerly called Nootka), there is a process that labializes certain consonants after a round vowel, and there is another process that delabializes them at the end of a syllable (Campbell 1973). When a consonant meets both of these conditions, what does it do? In derivational approaches, the rule that applies last gets to decide, as shown in (17). In OT, priority is determined by constraint ranking [see (18)].

(17) Nuuchahnulth /mˀoːq/ → [mˀoːq] with rules

Underlying	/mˀoːq/
q→qʷ /o___	mˀoːqʷ
qʷ→q /___.	mˀoːq
Surface	[mˀoːq]

(18) Nuuchahnulth /mˀoːq/ → [mˀoːq] in OT

/mˀoːq/	*qʷ.	*oq	Ident(round)
a. → mˀoːq		*	
b. mˀoːqʷ	*W	L	*W

Although standard approaches to OT have nothing like the derivations in (15) and (17), there are proposals to link several different OT grammars serially, with the output of one becoming the input to the next. This idea has obvious relevance in any situation where we are dealing with separate modules of the grammar, such as the syntax and the phonology. It has also been recruited in phonology to account for phenomena such as within-language differences between word and sentence phonology (McCarthy and Prince 1993: Appendix; Potter 1994; Kenstowicz 1995; Orgun 1996; Cohn and McCarthy 1994/1998; Hale Kissock and Reiss 1998; Bermúdez-Otero 1999; McCarthy 2000; Rubach 2000; Ito and Mester 2001, 2003a,b; Kiparsky 2003).

Suggestions for further reading. McCarthy (2007) includes an overview and bibliography of most work related to derivations in OT phonology. Many of the contributions to Hermans and van Oostendorp (1999) and Roca (1997) are also relevant.

6.8 GEN AND FAITHFULNESS BEYOND PHONOLOGY

In phonology, it's reasonably clear what GEN and the faithfulness constraints should look like. But what about syntax? In general, how do we go about establishing the nature of GEN and faithfulness in domains other than phonology?

There is another way of putting this question: what do OT's basic principles tell us about the necessary properties of GEN and the faithfulness part of CON in *any* empirical domain? One important idea is that the input and GEN together define the space of candidates that compete to be the surface realization of that input. This places lower and upper bounds on GEN's freedom of analysis, its capacity to create output candidates that are different from the input (see Section 3). If a hypothesized GEN is not free enough, then some inputs might have only observationally ungrammatical expressions among their output candidates. No adequate theory of GEN should ever do this, because every candidate set must contain *some* winner (Section 6.4). If a hypothesized GEN is too free, on the other hand, then some observationally grammatical realization of an input might lose to another, more harmonic candidate.[8] Clearly, assumptions about the nature of the input as well as assumptions about the nature of GEN will affect how these criteria work out in practice.

Another important idea is that any property of the input that GEN can alter – any property that can vary among the output candidates for a given input – must be protected by some faithfulness constraint if it is to affect the output of the grammar. The reason: apart from possible restrictions on GEN's freedom of analysis, faithfulness constraints are the only mechanism in OT for transmitting information from the input to the output.

From these statements, both of which derive from basic assumptions of the theory, it follows that a hypothesis about any one of the input, GEN, the candidate set, and the faithfulness constraints will go a long way toward determining all of the rest. Legendre, Smolensky, and Wilson (1998) show how this can be done in a real-life example, *wh* movement.

Suggestions for further reading. Blaho, Bye, and Krämer (forthcoming) is an anthology of work on GEN and freedom of analysis.

7. *Where to Go Next*

Readers whose interest in OT has been piqued by this brief overview should plan on reading Prince and Smolensky (1993/2004), especially parts I and III if their interests are not primarily phonological. This exercise is perhaps best preceded or accompanied by reading a more didactic treatment, such as Kager (1999), McCarthy (2002), or one of the other relatively accessible works cited at the end of Section 3.

The next step after that depends on the individual reader's interests. If they tend toward phonology, then the articles collected in McCarthy (2003b) are probably the best place to start. Two other useful anthologies, Lombardi (2001) and Féry and van de Vijver (2003), are focused on segmental and syllabic phonology, respectively. Readers of a syntactic bent could not do better than to consult two relatively recent anthologies of articles on OT syntax, Legendre, Grimshaw, and Vikner (2001) and Sells et al. (2001). In addition, there are now several anthologies on OT semantics and pragmatics (de Hoop and de Swart 1999; Blutner and Zeevat 2004; Blutner, de Hoop and Hendricks 2005), and one on historical linguistics (Holt 2003).

Finally, readers should be aware of two remarkable compendia of research on OT. One is the Rutgers Optimality Archive (ROA, http://roa.rutgers.edu). ROA, which was created by Alan Prince in 1993, is an electronic repository of 'work in, on, or about OT'. By the time this article is published, ROA will probably have its 1000th entry. The other notable compendium is the two-volume set *The Harmonic Mind*, edited by Paul Smolensky and Géraldine Legendre (2006). In almost 1200 pages, it collects the work of the editors and their collaborators on topics in language, human cognition, the mind, and the brain, much of it centered on Optimality Theory.

Short Biography

John J. McCarthy's research deals with various topics in phonology, prosodic morphology, and Optimality Theory. He is the author of *A Thematic Guide to Optimality Theory* (Cambridge, 2002) and *Hidden Generalizations: Phonological Opacity in Optimality Theory* (Equinox, 2007), and he edited *Optimality Theory in Phonology: A Reader* (Blackwell, 2004). His textbook *Doing Optimality Theory* will be published by Blackwell in 2008. His articles have appeared in the journals *Phonology*, *Natural Language and Linguistic Theory*, *Language*, and *Linguistic Inquiry*. He has taught courses on Optimality Theory at two Linguistic Society of America Summer Institutes (Cornell 1997 and MIT 2005). He received a Guggenheim Fellowship in 1992 and was elected to the American Academy of Arts and Sciences in 2005.

Notes

* Correspondence address: John J. McCarthy, Department of Linguistics, University of Massachusetts, 226 South College, 150 Hicks Way, Amherst, MA 01003-9274, USA. Email: jmccarthy@linguist.umass.edu.

[1] I am very grateful for comments and suggestions received from Kathryn Flack, Karen Jesney, Cathie Ringen, an anonymous reviewer, and the University of California San Diego review team (Eric Bakovic, Kathryn Davidson, Gabriel Doyle, J. Grant Loomis, Bozena Pajak, and Mohammad Salihie).

[2] In theory, EVAL could run out of constraints before the candidate set has been reduced to a single member. This can only happen if two or more candidates receive exactly the same number of violation marks from all of the constraints. In actual practice, that is unlikely to occur.

[3] Chomsky (1995: 380) criticizes faithfulness constraints on the grounds that identity between input and output is 'a principle that is virtually never satisfied'.

[4] See Karttunen (2006) for some discussion of this point.

[5] *n!*, pronounced 'n factorial', is the product $n * n-1 * n-2 * \ldots * 2$.

[6] Examples like /fi ɔpia/ → [fjɔpia] show that Emai tolerates vowel sequences word internally. Although Schaefer (1987) does not say how these sequences are syllabified, presumably at least some are heterosyllabic: [fjɔ.pi.a]. Onsetless syllables are therefore tolerated in word-medial position, although they are eliminated word intially. This is an indication that the high-ranking markedness constraint in Emai is specific to word-initial syllables. Flack (2007) has identified a number of languages that require onsets word initially but not medially, and this leads her to

distinguish between ONSET$_{Word}$ and ONSET$_{Syllable}$ constraints. Emai's tolerance for onsetless syllables phrase initially (e.g., [ɔlebe]) has a similar etiology.
[7] The Southern Min or Taiwanese tone circle has been cited as a potential counterexample to this claim. There is, however, a body of work arguing 'that Taiwanese tone sandhi is better viewed as a set of essentially arbitrary alternations between stored allomorphs' (Myers and Tsay 2002: 4). See Myers and Tsay for extensive references.
[8] This criterion must be used with care, because of the possibility of variable constraint ranking (Anttila 1997; Boersma and Hayes 2001 and others).

Works Cited

Aissen, Judith. 1999. Markedness and subject choice in Optimality Theory. Natural Language and Linguistic Theory 17.673–711.

Anderson, Stephen R. 1985. Phonology in the twentieth century: theories of rules and theories of representations. Chicago, IL: University of Chicago Press.

Anttila, Arto. 1997. Variation in Finnish phonology and morphology. Stanford University, Doctoral dissertation.

Archangeli, Diana. 1997. Optimality Theory: an introduction to linguistics in the 1990s. Optimality Theory: an overview, ed. by Diana Archangeli and D. Terence Langendoen, 1–32. Oxford, UK: Blackwell.

Bakovic, Eric and Edward Keer. 2001. Optionality and ineffability. Optimality-Theoretic syntax, ed. by Géraldine Legendre, Jane Grimshaw and Sten Vikner, 97–112. Cambridge, MA: MIT Press. (available at http://roa.rutgers.edu/view.php3?start=384.)

Becker, Michael. 2006. CCamelOT – An implementation of OT-CC's GEN and EVAL in Perl. Albuquerque, NM: Linguistic Society of America, 80th Annual Meeting.

Beckman, Jill. 1997. Positional faithfulness, positional neutralization, and Shona vowel harmony. Phonology 14.1–46.

——. 1999. Positional faithfulness: an Optimality Theoretic treatment of phonological asymmetries. New York: Garland. (1998 Doctoral dissertation, Amherst, MA: University of Massachusetts; available at http://roa.rutgers.edu/view.php3?start=234.)

Beckman, Jill N. and Catherine O. Ringen. 2004. Contrast and redundancy in OT. WCCFL 23 Proceedings, ed. by B. Schmeiser, V. Chand, A. Kelleher and A. Rodriguez, 101–14. Somerville, MA: Cascadilla Press.

Bermúdez-Otero, Ricardo. 1999. Constraint interaction in language change: quantity in English and Germanic, University of Manchester, Doctoral dissertation.

Blaho, Sylvia, Patrik Bye, and Martin Krämer. forthcoming 2007. Freedom of analysis? Berlin, Germany: Mouton de Gruyter.

Blutner, Reinhard, and Henk Zeevat. (eds.) 2004. Optimality Theory and pragmatics. Basingstoke, Hampshire, UK: Palgrave Macmillan.

Blutner, Reinhard, Helen de Hoop, and Petra Hendricks. (eds.) 2005. Optimal communication. Stanford, CA: CSLI Publications.

Boersma, Paul, and Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. Linguistic Inquiry 32.45–86. (available at http://roa.rutgers.edu/view.php3?start=348.)

Booij, Geert. 1995. The phonology of Dutch. Oxford, UK: Clarendon Press.

Brasoveanu, Adrian and Alan Prince. 2005. Ranking and necessity. New Brunswick, NJ: Rutgers University. (available at http://roa.rutgers.edu/view.php3?start=794.)

Bresnan, Joan. 2001a. Explaining morphosyntactic competition. The handbook of contemporary syntactic theory, ed. by Mark Baltin and Chris Collins, 11–44. Oxford, UK: Blackwell. (available at http://roa.rutgers.edu/view.php3?start=299.)

——. 2001b. The emergence of the unmarked pronoun. Optimality-Theoretic syntax, ed. by Géraldine Legendre, Jane Grimshaw and Sten Vikner, 113–42. Cambridge, MA: MIT Press. (available at http://roa.rutgers.edu/view.php3?start=179.)

Campbell, Lyle. 1973. Extrinsic order lives. Bloomington, IN: Indiana University Linguistics Club Publications.

Casali, Roderic F. 1996. Resolving Hiatus. UCLA, Doctoral dissertation. (available at http://roa.rutgers.edu/view.php3?start=215.)

Chomsky, Noam. 1957. Syntactic structures. The Hague, The Netherlands: Mouton.
——. 1968. Language and mind. New York: Harcourt Brace Jovanovich.
——. 1981. Lectures on government and binding. Dordrecht, The Netherlands: Foris.
——. 1995. The minimalist program. Cambridge, MA: MIT Press.
Chomsky, Noam and Morris Halle. 1968. The sound pattern of English. New York: Harper & Row.
Chomsky, Noam and Howard Lasnik. 1977. Filters and control. Linguistic Inquiry 8.425–504.
Cohn, Abigail and John J. McCarthy. 1994/1998. Alignment and parallelism in Indonesian phonology. Working papers of the Cornell phonetics laboratory 12.53–137. (available at http://roa.rutgers.edu/view.php3?start=25.)
de Hoop, Helen and Henriette de Swart. (eds.) 1999. Papers on optimality theoretic semantics. Utrecht, The Netherlands: Utrecht Institute of Linguistics/Onderzoeksinstituut voor Taal en Spraak.
Dinnsen, Daniel A. and Judith A. Gierut. (eds.) forthcoming 2007. Optimality Theory: phonological acquisition and disorders. London: Equinox Publishing.
Donegan, Patricia J. and David Stampe. 1979. The study of natural phonology. Current approaches to Phonological Theory, ed. by Daniel A. Dinnsen, 126–73. Bloomington, IN: Indiana University Press.
Eisner, Jason. 1997. Efficient generation in Primitive Optimality Theory. Philadelphia, PA: University of Pennsylvania. (available at http://roa.rutgers.edu/view.php3?start=206.)
Ellison, T. Mark. 1994. Phonological derivation in Optimality Theory. Proceedings of the 15th International Conference on Computational Linguistics (COLING), 1007–13. Kyoto. (available at http://roa.rutgers.edu/view.php3?start=75.)
Fanselow, Gisbert, Matthias Schlesewsky, Damir Cavar, and Reinhold Kliegl. 1999. Optimal Parsing, syntactic parsing preferences, and Optimality Theory. Potsdam, Germany: University of Potsdam. (available at http://roa.rutgers.edu/view.php3?start=367.)
Féry, Caroline and Ruben van de Vijver. 2003. The syllable in Optimality Theory. Cambridge, UK: Cambridge University Press.
Flack, Kathryn. 2007. The sources of phonological markedness. Amherst, MA: University of Massachusetts Amherst Doctoral dissertation.
Fodor, Jerry A., Thomas G. Bever, and Merrill Garrett. 1974. The psychology of language. New York: McGraw Hill.
Frank, Robert and Giorgio Satta. 1998. Optimality Theory and the computational complexity of constraint violability. Computational Linguistics 24.307–15. (available at http://roa.rutgers.edu/view.php3?start=228.)
Gerdemann, Dale and Gertjan van Noord. 2000. Approximation and exactness in finite state Optimality Theory. Tübingen and Groningen: University of Tübingen and University of Groningen. (available at http://arxiv.org/abs/cs/0006038v1).
Gibson, Edward and Kevin Broihier. 1998. Optimality Theory and human sentence processing. Is the best good enough? Optimality and competition in syntax, ed. by Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis and David Pesetsky, 157–91. Cambridge, MA: MIT Press.
Goldsmith, John. 1976a. Autosegmental phonology. MIT, Doctoral dissertation (Published by Garland Press, New York, 1979).
——. 1976b. An overview of autosegmental phonology. Linguistic Analysis 2.23–68.
——. 1993. Introduction. The last phonological rule: reflections on constraints and derivations, ed. by John Goldsmith, 1–20. Chicago, IL: University of Chicago Press.
Gouskova, Maria. 2003. Deriving economy: syncope in Optimality Theory. Amherst, MA: University of Massachusetts, Doctoral dissertation. (available at http://roa.rutgers.edu/view.php3?start=610.)
——. 2007. DEP: beyond epenthesis. Linguistic Inquiry 38.4.
Grimshaw, Jane. 1997. Projection, heads, and optimality. Linguistic Inquiry 28.373–422. (available at http://roa.rutgers.edu/view.php3?start=68.)
——. 2001. Optimal clitic positions and the lexicon in Romance clitic systems. Optimality-Theoretic syntax, ed. by Géraldine Legendre, Jane Grimshaw and Sten Vikner, 205–40. Cambridge, MA: MIT Press. (available at http://roa.rutgers.edu/view.php3?start=374.)
——. 2002. Economy of structure in OT. University of Massachusetts Occasional Papers in Linguistics 26: Papers in Optimality Theory II, ed. by Angela Carpenter, Andries Coetzee and Paul de Lacy, 81–120. Amherst, MA: GLSA. (available at http://roa.rutgers.edu/view.php3?start=434.)

Grimshaw, Jane and Vieri Samek-Lodovici. 1995. Optimal subjects. Papers in Optimality Theory, ed. by Jill Beckman, Laura Walsh Dickey and Suzanne Urbanczyk, 589–606. Amherst, MA: GLSA Publications.
——. 1998. Optimal subjects and subject universals. Is the best good enough? Optimality and competition in syntax, ed. by Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis and David Pesetsky, 193–219. Cambridge, MA: MIT Press.
Hale, Mark, Madelyn Kissock, and Charles Reiss. 1998. Output-output correspondence in Optimality Theory. The Proceedings of the West Coast Conference on Formal Linguistics 16, ed. by Emily Curtis, James Lyle and Gabriel Webster, 223–36. Stanford, CA: CSLI Publications. (available at http://roa.rutgers.edu/view.php3?start=202.)
Hammond, Michael. 1997. Parsing syllables: modeling OT computationally. Tucson, AZ: University of Arizona. (available at http://roa.rutgers.edu/view.php3?start=222.)
Hayes, Bruce. 1995. Metrical stress theory: principles and case studies. Chicago, IL: The University of Chicago Press.
Hermans, Ben and Marc van Oostendorp. (eds.) 1999. The derivational residue in phonological Optimality Theory. Amsterdam, The Netherlands: John Benjamins.
Holt, D. Eric (ed.) 2003. Optimality Theory and language change. Dordrecht, The Netherlands: Kluwer Academic.
Ito, Junko. 1989. A prosodic theory of epenthesis. Natural Language and Linguistic Theory 7.217–59.
Ito, Junko and Armin Mester. 2001. Structure preservation and stratal opacity in German. Segmental phonology in Optimality Theory, ed. by Linda Lombardi, 261–95. Cambridge, UK: Cambridge University Press.
——. 2003a. On the sources of opacity in OT: Coda processes in German. The syllable in Optimality Theory, ed. by Caroline Féry and Ruben van de Vijver, 271–303. Cambridge, UK: Cambridge University Press. (available at http://roa.rutgers.edu/view.php3?start=347.)
——. 2003b. Lexical and postlexical phonology in Optimality Theory: evidence from Japanese. Resolving conflicts in grammars: Optimality Theory in syntax, morphology, and phonology. Linguistische Berichte Sonderheft 11, ed. by Gisbert Fanselow and Caroline Féry, 183–207. (available at http://people.ucsc.edu/~mester/papers/2003_ito_mester_lexpostlex.pdf).
Kager, René. 1999. Optimality Theory. Cambridge, UK: Cambridge University Press.
Kager, René, Joe Pater, and Wim Zonneveld. (eds.) 2004. Constraints in phonological acquisition. Cambridge, UK: Cambridge University Press.
Karttunen, Lauri. 1998. The proper treatment of optimality in computational phonology. FSMNLP '98: Proceedings of the International Workshop on Finite State Methods in Natural Language Processing, 1–12. Ankara, Turkey: Bilkent University. (available at http://roa.rutgers.edu/view.php3?start=258.)
——. 2006. The insufficiency of paper-and-pencil linguistics: the case of Finnish prosody. Intelligent linguistic architectures: variations on themes by Ronald M. Kaplan, ed. by Miriam Butt, Mary Dalrymple and Tracy Holloway King, 287–300. Stanford, CA: CSLI Publications. (available at http://roa.rutgers.edu/view.php3?start=818.)
Kenstowicz, Michael. 1995. Cyclic vs. non-cyclic constraint evaluation. Phonology 12.397–436. (available at http://roa.rutgers.edu/view.php3?start=31.)
Kiparsky, Paul. 2003. Syllables and moras in Arabic. The syllable in Optimality Theory, ed. by Caroline Féry and Ruben van de Vijver, 147–82. Cambridge, UK: Cambridge University Press (available at http://www.stanford.edu/~kiparsky/Papers/syll.pdf).
Kisseberth, Charles. 1970. On the functional unity of phonological rules. Linguistic Inquiry 1.291–306.
Legendre, Géraldine. 2001. An introduction to Optimality Theory in syntax. Optimality-Theoretic syntax, ed. by Géraldine Legendre, Jane Grimshaw and Sten Vikner, 1–28. Cambridge, MA: MIT Press.
Legendre, Géraldine, Jane Grimshaw, and Sten Vikner. (eds.) 2001. Optimality-Theoretic syntax. Cambridge, MA: MIT Press.
Legendre, Géraldine, Paul Smolensky, and Colin Wilson. 1998. When is less more? Faithfulness and minimal links in wh-chains. Is the best good enough? Optimality and competition in syntax, ed. by Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis and David Pesetsky, 249–89. Cambridge, MA: MIT Press. (available at http://roa.rutgers.edu/view.php3?start=117.)

Legendre, Géraldine, Antonella Sorace, and Paul Smolensky. 2006. The Optimality Theory-Harmonic grammar connection. The harmonic mind: from neural computation to Optimality-Theoretic grammar, vol. 2: linguistic and philosophical implications, ed. by Paul Smolensky and Géraldine Legendre, 339–402. Cambridge, MA: MIT Press/Bradford Books.

Lombardi, Linda (ed.) 2001. Segmental phonology in Optimality Theory: constraints and representations. Cambridge, UK: Cambridge University Press.

McCarthy, John J. 2000. The prosody of phase in Rotuman. Natural Language and Linguistic Theory 18.147–97.

——. 2002. A thematic guide to Optimality Theory. Cambridge, UK: Cambridge University Press.

——. 2003a. Optimality Theory: an overview. Oxford international encyclopedia of linguistics, 2nd edn., ed. by William Frawley. Oxford, UK: Oxford University Press.

——. (ed.) 2003b. Optimality Theory in phonology: A reader. Malden, MA: Blackwell.

——. 2007. Hidden generalizations: phonological opacity in Optimality Theory. London: Equinox Publishing.

——. forthcoming 2008. Doing Optimality Theory. Malden, MA: Blackwell.

McCarthy, John J., and Alan Prince. 1993. Prosodic Morphology: Constraint Interaction and Satisfaction. New Brunswick, NJ: Rutgers University Center for Cognitive Science. (available at http://roa.rutgers.edu/view.php3?start=482.)

——. 1994. The emergence of the unmarked: optimality in prosodic morphology. Proceedings of the North East Linguistic Society 24, ed. by Mercè Gonzàlez, 333–79. Amherst, MA: GLSA Publications [available at http://roa.rutgers.edu/view.php3?start=13; excerpted in Optimality Theory in Phonology: A Reader, ed. by John J. McCarthy, Malden, MA and Oxford, Blackwell (2004)].

——. 1995. Faithfulness and Reduplicative Identity. University of Massachusetts Occasional Papers in Linguistics 18, ed. by Jill Beckman, Laura Walsh Dickey and Suzanne Urbanczyk, 249–384. Amherst, MA: GLSA Publications.

——. 1999. Faithfulness and identity in Prosodic Morphology. The Prosody-Morphology Interface, ed. by René Kager, Harry van der Hulst and Wim Zonneveld, 218–309. Cambridge: Cambridge University Press [excerpted in Optimality Theory in Phonology: A Reader, ed. by John J. McCarthy, Malden, MA and Oxford, Blackwell (2004)].

Moreton, Elliott. 2003. Non-computable functions in Optimality Theory. Optimality Theory in phonology: A Reader, ed. by John J. McCarthy, 141–63. Malden, MA: Blackwell. (available at http://roa.rutgers.edu/view.php3?start=364.)

Myers, James, and Jane Tsay. 2002. Neutralization in Taiwanese tone sandhi. Proceedings of the Berkeley Linguistics Society 22. 394–405. Taiwan: National Chung Cheng University (available at http://www.ccunix.ccu.edu.tw/~lngproc/MyersTsayTTSneut.pdf).

Orgun, C. Orhan. 1996. Sign-based morphology and phonology, with special attention to Optimality Theory. Berkeley, MA: University of California, Doctoral dissertation. (available at http://roa.rutgers.edu/view.php3?start=171.)

Orgun, C. Orhan, and Ronald Sprouse. 1999. From MParse to control: deriving ungrammaticality. Phonology 16.191–220. (available at http://roa.rutgers.edu/view.php3?start=224.)

Payne, David L. 1981. The phonology and morphology of Axininca Campa. Summer Institute of Linguistics Publications in Linguistics, 66. Dallas, TX: Summer Institute of Linguistics and University of Texas at Arlington.

Perlmutter, David. 1971. Deep and surface structure constraints in syntax. New York: Holt, Rinehart, & Winston.

Pesetsky, David. 1997. Optimality Theory and syntax: movement and pronunciation. Optimality Theory: an overview, ed. by Diana Archangeli and D. Terence Langendoen, 134–70. Oxford, UK: Blackwell.

Potter, Brian. 1994. Serial optimality in Mohawk prosody. Proceedings of the Thirtieth Annual Regional Meeting of the Chicago Linguistics Society, ed. by Katharine Beals et al., 347–61. Chicago, IL: Chicago Linguistics Society.

Prentice, D. J. 1971. The Murut languages of Sabah. Canberra, Australia: Australian National University.

Prince, Alan. 1983. Relating to the grid. Linguistic Inquiry 14.19–100.

——. 1997a. Endogenous constraints on Optimality Theory. Lecture notes for course presented at Summer Institute of the Linguistic Society of America, Cornell University, Ithaca, NY (available at http://ling.rutgers.edu/gamma/talks/insthdt1.pdf).

——. 1997b. Stringency and anti-Paninian hierarchies. Lecture notes for course presented at Summer Institute of the Linguistic Society of America, Cornell University, Ithaca, NY (available at http://ling.rutgers.edu/gamma/talks/insthdt2.pdf).

——. 2002. Arguing optimality. Papers in Optimality Theory II (= University of Massachusetts Occasional Papers 26), ed. by Angela Carpenter, Andries Coetzee and Paul de Lacy, 269–304. Amherst, MA: GLSA. (available at http://roa.rutgers.edu/view.php3?start=562.)

Prince, Alan and Paul Smolensky. 1993/2004. Optimality Theory: constraint interaction in generative grammar. Malden, MA, and Oxford, UK: Blackwell (revision of 1993 technical report, Rutgers University Center for Cognitive Science; available at http://roa.rutgers.edu/view.php3?start=537.)

——. 1997. Optimality: from neural networks to universal grammar. Science 275.1604–10.

——. 2003. Optimality Theory in phonology. Oxford international encyclopedia of linguistics, ed. by William Frawley, 212–222. Oxford, UK: Oxford University Press (available at http://ling.rutgers.edu/gamma/oiel.pdf).

Pulleyblank, Douglas. 1986. Tone in lexical phonology. Dordrecht, The Netherlands: D. Reidel.

Rice, Curt. 2003. Syllabic well-formedness in Norwegian imperatives. Nordlyd 31.372–384. (available at http://roa.rutgers.edu/view.php3?start=642 under title 'Dialectal variation in Norwegian imperatives').

——. 2005. Optimal gaps in optimal paradigms. Catalan Journal of Linguistics 4.155–70 (special issue on phonology in morphology edited by Maria-Rosa Lloret and Jesús Jiménez. Available at LingBuzz, http://ling.auf.net/buzzdocs/).

——. (ed.) forthcoming 2008. Modeling ungrammaticality in Optimality Theory. London: Equinox Publishing.

Raffelsiefen, Renate. 2004. Absolute ill-formedness and other morphophonological effects. Phonology 21.91–142.

Riggle, Jason. 2004. Generation, recognition, and learning in finite state optimality theory. UCLA, Doctoral dissertation (available at http://hum.uchicago.edu/~jriggle/riggle04_2.pdf.).

Roca, Iggy (ed.) 1997. Derivations and constraints in phonology. Oxford, UK: Oxford University Press.

Rubach, Jerzy. 2000. Glide and glottal stop insertion in Slavic languages: a DOT analysis. Linguistic Inquiry 31.271–317.

Samek-Lodovici, Vieri. 1996. Constraints on subjects: an Optimality Theoretic analysis. Rutgers University, Doctoral dissertation. (available at http://roa.rutgers.edu/view.php3?start=148.)

Schaefer, Ronald P. 1987. An initial orthography and lexicon for Emai. Bloomington, IN: Indiana University Linguistics Club Publications.

Sells, Peter, Joan Bresnan, Miriam Butt, and Tracy Holloway King. 2001. Formal and empirical issues in Optimality Theoretic syntax. Stanford, CA: CSLI Publications.

Smolensky, Paul and Géraldine Legendre. 2006. The harmonic mind: from neural computation to Optimality-Theoretic grammar. Cambridge, MA: MIT Press.

Smolensky, Paul, Géraldine Legendre, and Bruce Tesar. 2006. Optimality Theory: the structure, use, and acquisition of grammatical knowledge. The harmonic mind: from neural computation to Optimality-Theoretic grammar, vol. 1: Cognitive architecture, ed. by Paul Smolensky and Géraldine Legendre, 453–535. Cambridge, MA: MIT Press/Bradford Books.

Stampe, David. 1973. A dissertation on natural phonology. University of Chicago, Doctoral dissertation. (published by Garland, New York, 1979).

Tesar, Bruce. 1995. Computational Optimality Theory. University of Colorado, Doctoral dissertation. (available at http://roa.rutgers.edu/view.php3?start=90.)

——. 1998. Using the mutual inconsistency of structural descriptions to overcome ambiguity in language learning. Proceedings of the North East Linguistic Society 28, ed. by Pius N. Tamanji and Kiyomi Kusumoto, 469–83. Amherst, MA: GLSA Publications. (available at http://roa.rutgers.edu/view.php3?start=426.)

——. 1999. Robust interpretive parsing in metrical stress theory. The Proceedings of the West Coast Conference on Formal Linguistics 17, ed. by Kimary N. Shahin, Susan J. Blake and Eun-Sook Kim. Stanford, CA: CSLI Publications. (available at http://roa.rutgers.edu/view.php3?start=262.)

Tesar, Bruce and Alan Prince,. 2004. Using phonotactics to learn phonological alternations. CLS 39, Part II: The panels. Chicago, IL: Chicago Linguistic Society. (available at http://roa.rutgers.edu/view.php3?start=620.)

Tesar, Bruce and Paul Smolensky. 1998. Learnability in Optimality Theory. Linguistic Inquiry 29.229–68 [available at http://roa.rutgers.edu/view.php3?start=155; reprinted in Optimality Theory in Phonology: A Reader, ed. by John J. McCarthy, Malden, MA and Oxford, Blackwell (2004)].

——. 2000. Learnability in Optimality Theory. Cambridge, MA: MIT Press.

Tesar, Bruce, Jane Grimshaw, and Alan Prince. 1999. Linguistic and cognitive explanation in Optimality Theory. What is Cognitive Science? ed. by Ernest Lepore and Zenon Pylyshyn, 295–326. Oxford, UK: Blackwell.

Urbanczyk, Suzanne. 2006. Reduplicative form and the root-affix asymmetry. Natural Language and Linguistic Theory 24.179–240.

Wareham, Harold T. 1998. Systematic parametrized complexity analysis in computational phonology. University of Victoria, PhD dissertation. (available at http://roa.rutgers.edu/view.php3?start=318.)