Optimality Theory and the Generative Complexity of Constraint Violability

Robert Frank*
Johns Hopkins University

Giorgio Satta†
Università di Padova

It has been argued that rule-based phonological descriptions can uniformly be expressed as mappings carried out by finite-state transducers, and therefore fall within the class of rational relations. If this property of generative capacity is an empirically correct characterization of phonological mappings, it should hold of any sufficiently restrictive theory of phonology, whether it utilizes constraints or rewrite rules. In this paper, we investigate the conditions under which the phonological descriptions that are possible within the view of constraint interaction embodied in Optimality Theory (Prince and Smolensky 1993) remain within the class of rational relations. We show that this is true when GEN *is itself a rational relation, and each of the constraints distinguishes among finitely many regular sets of candidates.*

1. Introduction

Analyses within generative phonology have traditionally been stated in terms of systems of rewrite rules, which, when applied in the appropriate sequence, produce a surface form from an underlying representation. As first pointed out by Johnson (1972), the effects of phonological rewrite rules can be simulated using only finite-state machinery, with iterative application accomplished by sending the output from one transducer to the input of the next, a process that can be compiled out into a single transducer (Kaplan and Kay 1994).[1] Using this insight, a vast majority of computational implementations of phonological rule systems have been done using finite-state transducers or extensions thereof (Sproat 1992).

Recently, there has been a shift in much of the work on phonological theory, from systems of rules to sets of well-formedness constraints (Paradis 1988, Scobbie 1991, Prince and Smolensky 1993, Burzio 1994). This shift has, however, had relatively little impact upon computational work (but see Bird and Ellison 1994). In this paper, we begin an examination of the effects of the move from rule-based to constraint-based theories upon the generative properties of phonological theories. Specifically, we will focus our efforts on the issue of whether the widely adopted constraint-based view known as Optimality Theory (OT) may be instantiated in a finite-state transducer.[2] OT

* Department of Cognitive Science, 3400 N. Charles Street, Baltimore, MD 21218. E-mail: rfrank@cogsci.jhu.edu. This author is also affiliated with the Center for Language and Speech Processing, Johns Hopkins University.
† Dipartimento di Elettronica ed Informatica, Via Gradenigo 6/a, I-35131 Padova, Italy. E-mail: satta@dei.unipd.it. Part of the present research was done while this author was visiting the Center for Language and Speech Processing, Johns Hopkins University.
1 An alternative to composition of transducers involves running multiple rule transducers in parallel, producing so-called two-level phonological systems (Koskenniemi 1984). See Barton, Berwick, and Ristad (1987) for discussion of space and time complexity issues.
2 We are aware of two papers that study related matters. Ellison (1994) addresses the question of

raises a particularly interesting theoretical question in this context: it allows the specification of a ranking among the constraints and allows lower-ranked constraints to be violated in order for higher-ranked constraints to be satisfied. This violability property means that certain well-known computational techniques for imposing constraints are not directly applicable. Our study can be seen, therefore, as the beginnings of an investigation of the generative complexity of constraint ranking and violability. In this paper, we present a general formalization of OT that directly embodies that theory's notion of constraint violability. We then study the formal properties of one particular case of this general formalization in which the mapping from input to possible output forms, GEN, is representable as a finite-state transducer, and where each constraint is represented by means of some total function from strings to non-negative integers, with the requirement that the inverse image of every integer be a regular set. These two formal assumptions are sufficiently generous to allow us to capture most of the current phonological analyses within the OT framework that have been presented in the literature. We prove that the generative capacity of the resulting system does not exceed that of the class of finite-state transducers precisely when each constraint has a finite codomain, i.e., constraints may distinguish among only a finite set of equivalence classes of candidates. As will be discussed in Section 6, this result is optimal with respect to the finite codomain assumption, in the sense that dropping this assumption allows the representation of relations that cannot be implemented by means of a finite-state transducer (the latter fact has been shown to us by Markus Hiller, and will be discussed here). Before proceeding with the discussion of our result, however, we describe the rudiments of OT and introduce some technical notions.

2. Basics of OT

As in derivational systems, the general form of phonological computation in OT proceeds from an underlying representation (UR).[3] Such a UR is fed as input to the function GEN, which produces as output the set of all possible surface realizations (SRs) for this UR, called the **candidate set**. The notion of a possible SR, as realized in Prince and Smolensky (1993), is governed by the **containment condition**, requiring any SR output by GEN to include a representation of the UR as a (not necessarily contiguous) subpart. Thus, an SR must at a minimum include all of the structure that is specified in the UR, but may also include extra structure absent from the UR, called **epenthetic** structure. This is not to say that all parts of the input are necessarily pronounced at the surface. Rather, the analogue of "deletion" may occur by marking that part of the SR corresponding to the deleted material as **unparsed**, meaning that it is not visible to the phonetic interface.

The candidate set produced by GEN for any UR will in general be infinite, as there is no bound on the amount of epenthetic material that may be added to the UR to pro-

whether the constraint satisfaction problem for a specific input form can be compiled into a finite-state automaton. He provides an algorithm to produce a nondeterministic finite-state automaton that represents the set of winning candidates for any particular underlying form given finite-state representations of the input and the constraints. We are, however, interested in the more general question of whether the input–output mapping specified by OT for the class of inputs as a whole can be simulated with finite-state machinery. Another related study is that of Tesar (1995), who shows how the set of optimal output forms can be efficiently computed using a dynamic programming technique. Tesar does not, however, address the question of the generative complexity of the mappings his algorithm computes.

3 Length constraints prevent us from presenting a more comprehensive introduction to OT. For further discussion of the formal structure of the model and its empirical consequences, see Prince and Smolensky (1993) and references cited therein.

duce the SR. The core of the OT machinery is devoted to choosing among the members of this candidate set to determine which is the actual SR. To do this, OT imposes a set of well-formedness constraints on the elements of the candidate set. Note, however, that these constraints are not imposed conjunctively, meaning that the "winning" SR need not, and most often will not, satisfy them all. Instead, OT allows for the specification of a language-particular ranking among the constraints, reflecting their relative importance. The candidate SRs are evaluated with respect to the constraints in a number of stages. At each stage, the entire candidate set is subjected to one of the constraints, the stage at which a constraint is applied being determined by the specified constraint ranking.[4] There are two possible outcomes of such an evaluation. The first arises when some members of the candidate set violate the constraint, but others do not. In this case, the constraint permits us to distinguish among the members of the candidate set: those that do not satisfy the constraint are eliminated from the candidate set and are not considered in subsequent constraint evaluation. (Alternatively, if a constraint can be violated multiple times by a single SR, the relevant evaluation compares the number of violations incurred by each of the SRs in the candidate set. Candidates with the fewest violations are preferred and those with more violations are eliminated.) The second possible outcome from a constraint evaluation ensues when all of the members of the candidate set violate the constraint to the same degree, perhaps massively or perhaps not at all. When this happens, the constraint does not help us in narrowing down the candidate set. Hence, no candidates are eliminated from the candidate set and violations of the constraint do not block any of them from being considered further to be the actual SR. At the end of the last stage, i.e., when all constraints have been applied, what remains is precisely the subset of the candidate set that are the **optimal** satisfiers of the constraints under their ranking. This set of candidates, which will often contain only a single member under the system of constraints suggested by Prince and Smolensky (1993), is taken as the set of actual SRs for the original UR.

OT makes the strong assumption that the constraints used to evaluate the members of the candidate set are universal, and are therefore active in the phonology of every language. What varies from one language to another is the relative ranking of constraints. Thus, as soon as a commitment is made concerning the set of constraints, there is a concomitant commitment concerning the range of possible typological variation: every ordering of the constraints corresponds to a possible phonological system.

3. Formal Preliminaries

Before proceeding with our formalization of OT, it will be useful to review some formal notation. Given a finite alphabet Σ we denote by Σ^* the set of all strings over Σ, including the empty string ε, and we denote by 2^{Σ^*} the power set of Σ^*.

We assume that the reader is familiar with the notions of finite-state automaton, regular language, finite-state transducer, and rational relation; definitions and basic properties can be found in Gurari (1989). To recap briefly, a **finite-state transducer** is a finite-state automaton whose transitions are defined over the cross-product set $(\Sigma \cup \{\varepsilon\}) \times (\Delta \cup \{\varepsilon\})$, with Σ and Δ two (finite) alphabets. If we interpret Σ as the alphabet of input to the machine and Δ as the alphabet of output, each accepting

4 We note that there is nothing about the OT system that requires that candidates be evaluated in this serial manner. Instead, all of the constraints could be seen as being imposed in parallel, with the relative importance among violations being determined after the evaluation. From the perspective of specifying the abstract computation that is determined by the OT model, nothing hinges on this serial versus parallel distinction, so far as we can see.

computation of the transducer can be viewed as defining a mapping between a string in Σ^* and a string in Δ^*. Of course, the finite-state transducer may be nondeterministic, in which case a single input string may give rise to multiple outputs. Thus, every finite-state transducer can be associated with what is called a **rational relation**, a relation over $\Sigma^* \times \Delta^*$ containing all possible input–output pairs. A rational relation R can also be regarded as a function $[R]$ from Σ^* to 2^{Δ^*}, by taking $[R](u) = \{v \mid (u,v) \in R\}$ for each $u \in \Sigma^*$. We will use this latter representation of rational relations throughout our subsequent discussion.

4. A Model of OT

We are now in a position to present our formal model of the OT system. Let us denote as $\mathbf{N}$ the set of nonnegative integers.

Definition

An **optimality system** (OS) is a triple $G = (\Sigma, \textsc{gen}, C)$, where Σ is a finite alphabet, GEN is a relation over $\Sigma^* \times \Sigma^*$ and $C = \langle c_1, \ldots, c_p \rangle$, $p \geq 1$, is an ordered sequence of total functions from Σ^* to $\mathbf{N}$.

The basic idea underlying this definition is as follows: If w is a well-formed UR, $[\textsc{gen}](w)$ is the nonempty set of all associated SR, otherwise $[\textsc{gen}](w) = \emptyset$. Each function c in C represents some constraint of the grammar. For a given SR w, the nonnegative integer $c(w)$ is the "degree of violation" that w incurs with respect to the represented constraint. Given a set of candidates S, we are interested in the subset of S that violates c to the least degree, i.e., whose value under the function c is lowest. To facilitate reference to this subset, we define

$$\text{argmin}_c\{S\} = \{w \mid w \in S,\ c(w) = \min\{c(w') \mid w' \in S\}\}.$$

We can now define the map an OS induces. We do this in stages, each one representing the evaluation of the candidates according to one of the constraints. For each $w \in \Sigma^*$ and for $0 \leq i \leq p$ we define a function from Σ^* to 2^{Σ^*}:

$$\text{OT}^i_G(w) = \begin{cases} [\textsc{gen}](w) & \text{if } i = 0; \\ \text{OT}^{i-1}_G(w) & \text{if } i \geq 1 \text{ and } \text{argmin}_{c_i}\{\text{OT}^{i-1}_G(w)\} = \text{OT}^{i-1}_G(w); \\ \text{argmin}_{c_i}\{\text{OT}^{i-1}_G(w)\} & \text{if } i \geq 1 \text{ and } \text{argmin}_{c_i}\{\text{OT}^{i-1}_G(w)\} \neq \text{OT}^{i-1}_G(w). \end{cases}$$

Function OT^p_G is called the **optimality function** associated with G, and is simply denoted as OT_G. We drop the subscript when there is no ambiguity.

The question of the expressive power of OT can now be stated precisely: what is the generative capacity of the class of optimality functions? The answer to this question depends, of course, upon the character of the functions that serve as GEN and the constraints. Though we will not make any substantive empirical claims about these functions, we will make a number of specific assumptions concerning their formal nature. Regarding GEN, we assume that the mapping from the UR to the candidate set is specifiable in terms of a finite-state transducer, that is to say, we will consider only OSs for which GEN is a rational relation (viewing rational relations as functions, as specified in the previous section). Since the question that we focus on in this research is that of determining whether the class of mappings specifiable in OT is beyond the formal power of finite-state transducers, allowing GEN to be beyond the power of a finite-state transducer would decide the question by *fiat*.[5] In addition, we assume that each constraint c in C is **regular** in that it satisfies the following requirement: For each $k \in \mathbf{N}$, the set $\{w \mid w \in \Sigma^*,\ c(w) = k\}$ (i.e., the inverse image of k under c) is a regular language. In other words, this requires that the set of candidates that violate a given constraint to any particular level must be regular. The choice of regular constraints is for reasons essentially identical to those that motivated the use of rational relations for GEN.

It turns out that nearly all of the constraints that have been proposed in the OT phonological literature are regular in this sense. The reason for this is that OT constraints have tended to take the form of local conditions on the well-formedness of phonological representations, where local means bounded in size. Because of this restriction, we can characterize all possible violations of a given constraint c through a finite set of configurations V_c. More precisely, a phonological representations w attests as many violations of c as the number of occurrences of strings in V_c appearing as substrings of w. Since V_c is finite, it can be represented through some regular expression. Under the standard assumption that phonological representations are not structurally recursive, but rather are combined using essentially iterated concatenation, we can use well-known algebraic properties of regular languages (see for instance Kaplan and Kay 1994) to show that c is regular. (See Tesar 1995 for further discussion of a related notion of locality in constraints.)

5. OT as a Rational Relation

This section presents the main result of this paper. We show that OSs of the sort outlined in the last section can be implemented through finite-state transducers so long as each constraint of the system satisfies one additional restriction: that it have a finite codomain, meaning that it distinguishes among only a finite set of equivalence classes of candidates. We start with some properties of the class of rational relations that will be needed later (proofs of these properties can be found for instance in Gurari 1989). Let R be a rational relation. The **left projection** of R is the language $\text{Left}(R) = \{u \mid (u,v) \in R\}$. Symmetrically, the **right projection** is the language $\text{Right}(R) = \{v \mid (u,v) \in R\}$. It is well known that Left(R) and Right(R) are both regular languages. If R' is a rational relation, the composition of R and R', defined as $R \circ R' = \{(u,v) \mid (u,w) \in R,\ (w,v) \in R',\ \text{for some } w\}$, is still a rational relation.

Let L be a regular language. We define the **left restriction** of R to L as the relation $\text{Lrst}(R,L) = \{(u,v) \mid (u,v) \in R,\ u \in L\}$. Symmetrically, $\text{Rrst}(R,L) = \{(u,v) \mid (u,v) \in R,\ v \in L\}$ is the **right restriction** of R to L. Both Lrst(R,L) and Rrst(R,L) are rational relations. The idea underlying a proof of this fact is to compose R (to the left or to the right) with the identity relation $\{(w,w) \mid w \in L\}$, which is rational.

Let $G = (\Sigma, \text{GEN}, C)$ be an OS. We start the presentation of our result by restricting our attention to constraints having codomain of size two, that is, each c_i in C is a total function from Σ^* to $\{0,1\}$ such that both the set $L(c_i) = \{w \mid w \in \Sigma^*,\ c_i(w) = 0\}$ and its complement are regular. Recall that $L(c_i)$ denotes the language of all strings in Σ^* that satisfy the constraint of the grammar represented by c_i, and its complement, the strings

5 We recognize that this assumption, while plausible for phonological representations, is perhaps less so for syntactic representations. Further, as a reviewer points out, recent developments of OT in the domain of reduplication phenomena (McCarthy and Prince 1995), which assume that GEN produces a correspondence relation between the UR and SR, might constitute a phonological case in which GEN is not a rational relation. If well-formedness conditions on this correspondence relation are guaranteed only by the constraints, however, GEN could remain rational, though the constraints would no doubt cease to be expressible as regular languages.

mapped to 1 by c_i, includes all strings that violate it. Thus, such c_is correspond to constraints that can distinguish only between complete satisfaction and violation. Using the above restriction, we can reformulate the definition of OT^i reported in Section 4:

$$\mathrm{OT}^i(w) = \begin{cases} [\textsc{gen}](w) & \text{if } i = 0; \\ \mathrm{OT}^{i-1}(w) & \text{if } i \geq 1 \text{ and } \mathrm{OT}^{i-1}(w) \cap L(c_i) = \emptyset; \\ \mathrm{OT}^{i-1}(w) \cap L(c_i) & \text{if } i \geq 1 \text{ and } \mathrm{OT}^{i-1}(w) \cap L(c_i) \neq \emptyset. \end{cases} \quad (1)$$

Note that the case where all candidates in $\mathrm{OT}^{i-1}(w)$ satisfy constraint c_i falls under the second clause of the definition in Section 4, but under the third clause of (1). However, this case is treated in the same way in both definitions, since $\mathrm{OT}^{i-1}(w) = \mathrm{OT}^{i-1}(w) \cap L(c_i)$ if $\mathrm{OT}^{i-1}(w) \subseteq L(c_i)$. We are now ready to prove a technical lemma.

Lemma 1
Let $G = (\Sigma, \textsc{gen}, C)$ be an OS such that GEN is a rational relation and each constraint in C is regular and has co-domain of size two. Then OT_G is a rational relation.

Let us start with the basic idea underlying the proof of this lemma. Assume that for $i \geq 1$ we have already been able to represent OT^{i-1} by means of a rational relation R. Consider some UR w and the set of associated candidate SRs that are optimal with respect to OT^{i-1}, that is, the set $\mathrm{OT}^{i-1}(w) = [R](w)$. To compute the strings in this set that are optimal with respect to c_i, we must perform what amounts to a "conditional intersection" with the regular language $L(c_i)$, as determined by (1). That is, we check if there are candidates from $[R](w)$ that are also compatible with c_i, i.e., that are members of $L(c_i)$. If there are some some, we eliminate any nonsatisfying candidates by intersecting $[R](w)$ with $L(c_i)$ (third condition in [1]). However, if no such candidates remain, we do nothing to the set of candidates from OT^{i-1} (second condition in [1]). As shown in the proof below, it turns out that this can be done by partitioning the left projection of relation R into two regular languages. This results in the "splitting" of R into two relations, one of which must be "refined" by taking its right restriction to language $L(c_i)$. The union of the two resulting relations is then the desired representation of OT^i. Putting these ideas together, we are now ready to present a formal proof.

Proof
We show that OT^i is a rational relation for $0 \leq i \leq p$. We proceed by induction on i. For $i = 0$, the claim directly follows from our assumptions about GEN. Let $1 \leq i \leq p$. From the inductive hypothesis, there exists a rational relation R such that $[R] = \mathrm{OT}^{i-1}$. Since $L(c_i)$ is a regular language, from an already mentioned property it follows that:

$$R_1 = \mathrm{Rrst}(R, L(c_i))$$

is a rational relation as well. Function $[R_1]$ associates a UR to the set of SRs that are optimal up to constraint c_{i-1} and that also satisfy c_i, the latter being the effect of the right restriction operator. Since R_1 is rational, we have that $L_1 = \mathrm{Left}(R_1)$, the set of URs for which function $[R_1]$ results in some non-empty set, is a regular language. By a well-known closure property of regular languages, the complement of L_1, $\overline{L_1} = \Sigma^* - L_1$, is a regular language as well. Note that, for each UR in $\overline{L_1}$, no associated SR is both optimal up to constraint c_{i-1} and satisfies c_i. It then follows, by an already mentioned property, that:

$$R_2 = \mathrm{Lrst}(R, \overline{L_1})$$

is a rational relation. Note that function $[R_2]$ computes optimality up to constraint c_{i-1}, but only over those URs whose optimal satisfiers do not satisfy c_i. It is not difficult to see from an inspection of (1) that $\mathrm{OT}^i = [R_1 \cup R_2]$. Then the statement of the lemma follows from the fact that the class of rational relations is closed under finite union (see for instance Gurari 1989). □

The result in the above lemma can be extended to regular constraints having arbitrarily large finite codomain, corresponding to constraints that rank candidates along some finite-valued scale. This is done using a construction, first suggested in Ellison (1994), which, expressed intuitively, replaces any such constraint function by a finite number of constraint functions having codomain of size two. More formally, assume constraint c has codomain $\{0, 1, \ldots, k\}$, $k > 1$. We introduce new constraints $\langle c, i \rangle$, $1 \leq i \leq k$, defined as follows: For each $1 \leq i \leq k$ and $w \in \Sigma^*$, we let $\langle c, i \rangle(w) = 0$ if $c(w) < i$, $\langle c, i \rangle(w) = 1$ if $c(w) \geq i$. Each $\langle c, i \rangle$ has codomain of size two. Since the class of regular languages is closed under finite union, if c is regular then each $\langle c, i \rangle$ is regular.

We can finally state our main result, which directly follows from the above discussion and from Lemma 1.

Theorem 1
Let $G = (\Sigma, \text{GEN}, C)$ be an OS such that GEN is a rational relation and each constraint in C is regular and has a finite codomain. Then OT_G is a rational relation.

6. Discussion

We have shown that when GEN is a rational relation and the constraints have a finite codomain, constraint ranking as defined by OT yields a system whose generative capacity does not exceed that of rational relations. Because of the nature of the construction in the proof of Lemma 1 (specifically the union of the relations R_1 and R_2 at each stage in the iteration), the finite-state transducer that is built crucially exploits transition nondeterminism. We note, however, that any finite-state transducer used to implement an OS will in any case need to be nondeterministic, since in general OT can pair more than one SR with a given UR.[6]

As we have mentioned above, our result tolerates only so-called binary and multivalued constraints, constraints that rank the candidates along some finite-valued scale. A linguistic example of such a multivalued constraint is Prince and Smolensky's HNUC, which rates the goodness of a segment serving as a syllabic nucleus, the rating being determined by the position of the segment along the finitely partitioned sonority hierarchy. Yet, this formal power is not sufficient to express the greater proportion of phonological analyses that have been given in the OT framework. In particular, it is usually assumed that constraints can be violated an arbitrary number of times by a single form, and that differences at any level of violation are grammatically significant. For example, even in the simple system of syllable structure constraints discussed in Prince and Smolensky (1993, Section 6), the computation of optimality for certain

6 It is interesting to note that this potential for nondeterminism is not exploited under many of the systems of constraints that have actually been proposed by OT practitioners. For example, the existence of families of constraints requiring the alignment of particular morphemes with a certain boundaries in an SR, members of the family of so-called **generalized alignment** constraints (McCarthy and Prince 1993), will often have the effect of linearly ordering all SRs according to their optimality, thereby yielding a single SR for each UR.

very long forms might require us to distinguish between 300 and 301 violations of the PARSE constraint. Consequently, it is a question of significant interest whether our result extends to the case of such gradient constraints, or in more formal terms, whether OT_G remains a rational relation when the (regular) constraints of the system can have an unbounded codomain.

It turns out that this is not true in the general case. The following example (due to P. Smolensky, after an idea of M. Hiller who first proved this separation result) shows this fact using only a single constraint:

$$\begin{aligned} \Sigma &= \{a,b\}, \\ \text{GEN} &= \{(a^n b^m, a^n b^m) \mid n,m \in \mathbf{N}\} \cup \{(a^n b^m, b^n a^m) \mid n,m \in \mathbf{N}\}, \\ c(w) &= \#_a(w), \end{aligned}$$

where $\#_a(w)$ denotes the number of occurrences of a within w. (Constraint c can be seen as a prohibition against the occurrence of the letter a in an SR.) Clearly GEN is a rational relation and c satisfies our previous assumptions. It is not difficult to see that this system is associated with a function OT_G such that a string of the form $a^n b^m$ is mapped to the singleton $\{a^n b^m\}$ if $n < m$, to the singleton $\{b^n a^m\}$ if $m < n$, and to the set $\{a^n b^m, b^n a^m\}$ when $n = m$. The relation R that realizes such a function is not rational, since its right restriction to the regular language $\{a^n b^m \mid n,m \in \mathbf{N}\}$ does not have a regular left projection, namely $\{a^n b^m \mid n \leq m\}$. This fact shows that the result in Theorem 1 is optimal with respect to the finite codomain hypothesis, that is to say, no weaker assumption concerning the nature of the constraints will suffice to keep the generative capacity of mappings defined by OSs within that of rational relations. It remains an open problem to characterize precisely the generative capacity of systems with gradient constraints, as well as that of OSs with other assumptions about the formal power of GEN and the constraints.

Finally, it is useful to recall the empirical argument given in Karttunen (1993) that attested phonological processes mediating between UR and SR can be modeled by a finite-state transducer. Though this argument was given in the context of a different conception of phonological derivation, the conclusion, if correct, is general. That is, whether the relation between UR and SR is best characterized in terms of rewriting sequences or OT optimizations, Karttunen's argument suggests that the generative complexity of the resulting mapping need be no greater than that of rational translations. If this empirical argument is on the right track, our results diagnose a formal deficiency with the OT formal system, namely that it is too rich in generative capacity. Our results also suggest a cure, however: constraints should be limited in the number of distinctions they can make in levels of violation. We suspect that following this regimen will necessitate a shift in the type of optimization carried out in OT, from global optimization over arbitrarily large representations to local optimization over structural domains of bounded complexity (where only a bounded number of violations can possible occur). Following the empirical and formal implications of this move would go well beyond the scope of the present work, so we leave this for the future.

Acknowledgments
We wish to thank Markus Hiller, Martin Kay, Mehryar Mohri, and Paul Smolensky for helpful discussions related to this work. We are also indebted to three anonymous referees for comments that have helped to significantly improve our presentation.

References

Barton, G. Edward, Robert C. Berwick, and

Eric Sven Ristad. 1987. *Computational Complexity and Natural Language.* MIT Press, Cambridge, MA.
Bird, Steven and T. Mark Ellison. 1994. One-level phonology: Autosegmental representations and rules as finite automata. *Computational Linguistics,* 20(1):55–90.
Burzio, Luigi. 1994. *Principles of English Stress.* Cambridge University Press, Cambridge.
Ellison, T. Mark. 1994. Phonological derivation in optimality theory. In *Proceedings of the 15th International Conference on Computational Linguistics,* pages 1007–1013.
Gurari, Eitan. 1989. *An Introduction to the Theory of Computation.* Computer Science Press, New York, NY.
Johnson, C. Douglas. 1972. *Formal Aspects of Phonological Description.* Mouton, The Hague.
Kaplan, Ronald M. and Martin Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics,* 20(3):331–378. Written in 1980.
Karttunen, Lauri. 1993. Finite-state constraints. In John Goldsmith, editor, *The Last Phonological Rule.* University of Chicago Press, Chicago, pages 173–194.
Koskenniemi, Kimmo. 1984. A general computational model for word-form recognition and production. In *Proceedings of the 10th International Conference on Computational Linguistics,* pages 178–181.
McCarthy, John and Alan Prince. 1993. Generalized alignment. In Geert Booij and Jaap van Marle, editors, *Yearbook of Morphology 1993.* Kluwer, Dordrecht, pages 79–153.
McCarthy, John and Alan Prince. 1995. Faithfulness and reduplicative identity. In *UMOP 18: Papers in Optimality Theory.* Graduate Linguistics Students Association, University of Massachusetts, Amherst, pages 249–384.
Paradis, Carole. 1988. On constraints and repair strategies. *The Linguistic Review,* 6(1):71–97.
Prince, Alan and Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. Manuscript, Rutgers University and University of Colorado, Boulder.
Scobbie, James. 1991. *Attribute Value Phonology.* Ph.D. thesis, University of Edinburgh.
Sproat, Richard. 1992. *Morphology and Computation.* MIT Press, Cambridge, MA.
Tesar, Bruce. 1995. *Computational Optimality Theory.* Ph.D. thesis, University of Colorado, Boulder.