Introduction
The Emergence of Constraints in Generative Phonology and a Comparison of Three Current Constraint-Based Models

DARLENE LACHARITÉ
and
CAROLE PARADIS
Université Laval

Constraints, which go also by the names “conditions”, “filters”, and less transparently, “principles” and “parameters” are currently opposed to (contextual and non-contextual) arbitrary rules, and are argued to be more explanatory in the scientific sense that they can: 1) reduce the number of sources and/or causes of a given phenomenon; 2) link apparently unrelated facts; and 3) make more predictions, if formulated adequately and related to universal grammar (UG). Although the notion of phonological constraints was not born with generative phonology (see, e.g., Hockett 1958:282), it is within that framework that they have begun to receive more formal treatment (Paradis and Nikiema 1993:45).

The first part of this article1 provides a brief retrospective on the emergence of constraints within the generative framework.2 More precisely, we

1 We are very grateful to Keren Rice who sent prompt and detailed comments on this article. We are also indebted to our research assistants, Caroline Lebel and Yvan Rose, for their reactions and invaluable suggestions. This paper and the whole volume has benefited from discussions with several people. In particular, we would like to thank Robert Kirchner, John McCarthy, Emmanuel Nikiema and Henrietta Hung. Finally, we thank William Cowan and Christina Thiele for meticulous editorial attention to this and other articles in the volume. We acknowledge SSHRC grant 410-90-0575 and FCAR grant 90-NC-6383 to Carole Paradis, and financial support from the Faculté des lettres at Université Laval.

2 Paradis and Prunet (this volume) also provide a brief retrospective on constraints, but for the purpose of arguing against morpheme structure constraints.
will briefly consider the influence, concerns and approaches of Halle (1959), Stanley (1967), Chomsky and Halle (1968; henceforth SPE), Postal (1968), Kisseberth (1970), Natural Generative Phonology, and finally Multilinear Phonology. Even though the earlier influences often go unacknowledged, they have nonetheless contributed substantially to current constraint-based approaches, including the three that are represented in this volume: Declarative Phonology (DP), Optimality Theory (OT) and the Theory of Constraints and Repair Strategies (TCRS).

In Section 2, DP, OT, and TCRS will be compared and contrasted on the basis of their answers to the following questions: What is a constraint (Section 2.1)? Can constraints conflict (Section 2.2)? How are constraint conflicts resolved (Section 2.3)? Can constraints be violated (Section 2.4)? Are constraint violations repaired (Section 2.5)? How are surface exceptions accounted for (Section 2.6)? The conclusion is presented in Section 3. We will see that the three theories share fundamental similarities in spite of crucial differences. Along the way, we will touch on issues that are raised by the articles in this volume. We hope that, in providing a brief retrospective on the emergence of constraints, and a clearer picture of the points of consensus and difference among the three current constraint-based models, this article will stimulate debate and lead to further clarification and development of the constraint-based approach.

1. The Emergence of Constraints in Generative Phonology

1.1. Halle (1959)
Halle (1959) proposes a redundancy-free underlying representation that is sufficiently abstract to permit the derivation of all associated variants of the same root.\(^3\) The relationship between a surface form and this abstract underlying form is partly encoded in morpheme structure rules, which specify, and hence constrain, what and how redundant information is added, to prevent the derivation of unattested forms. A major problem with this rule-based approach is that, although it may be descriptively adequate, it is overly powerful and lacks predictive power, problems identified by Stanley (1967).

---

\(^{3}\)The underlying form in generative phonology is not simply the surface form minus the redundant information as in functionalism, but a more abstract form, although there is no general mechanism for establishing this abstract form. That has to be done on a case-by-case basis (see Halle 1959:57; Kenstowice and Kisseberth 1977).
of labour between, and the relative explanatory power of, the different constraint types. This task was taken up by Natural Generative Phonologists.

1.6. Natural Generative Phonology
Theorists working within the framework of Natural Generative Phonology not only accept, but grant a central role to post-morphemic constraints (Surface Phonetic Constraints) and recognize them as the principle source of phonological alternations. Shibatani (1973) presents strong arguments for post-morphemic constraints, showing, first, that there are cases where a given constraint seems to be both morphemic and post-morphemic, and second, that a post-morphemic constraint cannot always be replaced by a morphemic one (an MSC). Shibatani's treatment paves the way for Hooper's (1979) complete rejection of an underlying representation, and hence MSCs, in favour of surface constraints. Hooper argues that speakers' rules are formulated directly and exclusively from surface forms and that the work of MSCs can be equally well handled by syllabic constraints.

1.7. Multilinear Generative Phonology
Natural Generative Phonology has had a marked influence on multilinear phonology, notably the addition of a syllabic tier to the phonological representation. However, the emergence of multilinear phonology (which refers to the several tiers, outside the traditional segmental one, that now make up a phonological representation, as shown in 2) is commonly linked to Goldsmith's (1976) autosegmental theory and Kahn's (1976) syllabic theory. Goldsmith proposes dissociating tones and certain vocalic harmony features ([round] and [back]) from the feature matrix, to account for their autosegmental behaviour, an approach quickly adopted by those working on prosody.

(2) Levels of Representation in Multilinear Phonology:

```
metrical tier
|-- syllabic tier
   |-- timing tier
   |    |-- tonal tier
   |    |    |-- autosegmental tier
   |    |    |    |-- segmental tier
```
Of relevance here is the fact that the explosion of refinements in phonological representations has led to an increasing need for constraints to regulate the relations between units both on the same and on different representational tiers. For example, Goldsmith proposes the No Line Crossing Convention, which prohibits association lines linking units belonging to two different tiers from crossing each other as shown in (3).

(3) No Line Crossing Convention:
```
 X  X  X
 /   /   /
[F] [F] [F]
```

Since multilinear representations are found at the post-morphemic level, it is clear that constraints, such as the No Line Crossing Convention, along with numerous other conditions that apply to each representational tier, must be active post-morphemically.

The widespread acceptance of de facto post-morphemic constraints such as the No Line Crossing Convention would suggest that earlier debates over the nature and respective roles of rules and (both morphemic and post-morphemic) constraints had been settled. However, the issues have not been settled, so much as forgotten. Post-morphemic constraints have been proposed without links to the debates of the 60s, namely without regard for the redundancy problem discussed by Postal (1968), Hooper (1976) and Clayton (1976), among others. The same charge can be leveled at recent discussion over the function of rules. For example, Yip (1988), whose article relies heavily on the conspiratorial function of rules and surface adherence of outputs to constraints, in addition to expounding the advantages of context-free rules (the context being provided by the constraint served by the rules), does not refer to Kisseberth (1970), who drew our attention to similar points. Actually, the different theories of underspecification in multilinear phonology (see, e.g., Kiparsky 1982, Archangeli 1984, Pulleyblank 1986) have generally given more consideration to the models of the 60s than the different phonologists who proposed constraints during the course of the 70s and 80s (e.g., Goldsmith 1976; McCarthy 1986; Hume 1988; etc.). Notably, underspecification accounts post-morphemically, rather than via MSCs, for the addition of redundant features, forcing underspecification theorists to address the problem raised by Stanley (1967) concerning the potentially ternary use of binary features (see, e.g., the distinctiveness constraints, a solution proposed by Pulleyblank 1986:135 to this problem; see also Dresher 1985; Archangeli 1988:124; and Paradis and Prunet 1991:23–24 for a discussion).

In general, the end of the 80s was marked by an implicit reconsideration of Postal's redundancy problem and, more important yet, the question of how constraints fit into the phonology (see Kiparsky 1982 for proposals within the framework of lexical phonology). Already by the beginning of the 80s, some authors were asking whether we need arbitrary (contextual or non-contextual) rules in addition to context-free phonological rules motivated by constraints. Certain researchers squarely proposed eliminating all arbitrary phonological rules in favour of constraints (see Kaye et al 1985:305). However, Kaye et al's model is relatively informal in its use of constraints. For instance, it does not address some of the issues that are of fundamental concern to constraint-based theories, such as precisely how constraints are to take over from rules. Constraints in that view both specify the problem and effect the change in a way that either allows each constraint to handle only a single task at a time, without any possible account for the conspiracy effect, or to perform many different changes in an unpredictable fashion. There seems little advantage here in exchanging arbitrary rules for constraints. The upshot is that, while Kaye et al see the relationship between constraints and surface alternations, they do not formally account for it.

1.8. Summary

In this section, we have attempted to provide a brief family history of current constraint use and the surrounding issues. Clearly, this retrospective is incomplete (for a lengthier retrospective, see Paradis and Nikiema 1993), but we hope that it will contribute to a better understanding of what is really involved in a notion which is of particular importance in multilinear phonology. Too often, we tend to regard ideas as entirely new, as a bolt out of the blue, when what is new is our recognition of their crucial theoretical importance.


This section highlights important points of consensus as well as fundamental differences among DP, OT and TCRS. DP is distinguished from both OT and TCRS chiefly by virtue of being a monosstral framework using hard (i.e., inviolable) language-specific constraints, which also function as redundancy rules. It is represented in this volume by Scobbie's article; more detailed...
discussion of the framework can be found in Bird et al. (1999), Coleman (1993) and Scobbie (1991, 1992). In contrast, OT relies on universal, soft (i.e., violable) constraints that are prioritized on a language-specific basis. In OT, constraints are more passive than in either DP, where they actively construct the output form, or in TCRS where, as we shall see, constraints direct the construction of the output form. Constraints in OT select the output form from a field of candidates. They have no role in shaping the output, in contrast to both DP and TCRS, where constraints, directly or indirectly, constitute the force behind a phonological structure or change. OT is represented in this volume in the article by McCarthy and that by Itō and Mester; the framework is fully described in Prince and Smolensky (1993); see also McCarthy and Prince (1993). The third constraint-based framework represented in this volume is TCRS, laid out in Paradis (1988a, 1988b, 1990a, 1990b, 1993b) and Paradis and Prunet (1988), etc. TCRS is one of the first theories to replace arbitrary rules with motivated ones (repair strategies) in the context of a more formal constraint-based theory, and to provide a principled relationship between constraints and surface alternations. Unlike either DP or OT, TCRS has repair strategies that enforce conformity to universal and language-specific constraints. The articles by Béland et al.; LaCharité; Paradis; and Paradis and Prunet, in this volume, present recent developments and applications of the TCRS program.

Apart from Kaye et al. (1983), briefly discussed in Section 1.7, there are, to our knowledge, two other constraint-based theories. The first is Generative Phonotactics (GP) (Singh 1985, 1987, 1992), to which TCRS is sometimes compared—though, in fact, GP and TCRS essentially share only the notion of “repair strategy”, which was already present (albeit informally) in Kisseberth (1970) (see Paradis 1988a for a discussion of the differences between GP and TCRS). The other constraint-based theory is Harmonic Phonology, a highly formalized approach and one of the first to apply to phonology ideas basic to parallel distributed processing (McClelland and Rumelhart 1987). Harmonic Phonology is represented in a sizeable body of research (see Goldsmith 1991, 1993; Goldsmith and Larson 1992, etc.). However, these theories will not be discussed any further here: due to space limitations, this introduction is restricted to the presentation of the theories represented in this volume.

2.1. What is a Constraint?

DP, OT and TCRS all do away with *SPE*-type contextual rules in favour of constraints. Though all three theories implicitly or explicitly agree that a first in a jigsaw puzzle does not change the resulting picture. In DP, every word-building process and/or component is a constraint (see Scobbie, this volume) and the output of one constraint is never the input to another.

In DP, constraints are considered to be language-specific, positive or implicational statements of well-formedness that combine in a generally unordered fashion to construct a phonological form. In other words, the surface form is a composite of the relevant constraints of the language. Each constraint is a partial description that either licenses, contributes information to or is a piece of a well-formed surface structure (Coleman 1993:18). The lexical entry itself is considered to be one of the constraints that creates the surface form (Scobbie, this volume). As already mentioned, DP's view might be compared to a jigsaw puzzle: each constraint, including the lexical entry, is one piece of the puzzle; the completed picture represents the surface-phonological form. Whether one inserts one or another piece first has no effect on the result.

DP is considered to include both more and less marked language-specific constraints. However, despite this appeal to markedness, DP, in contrast with TCRS to be discussed further on, makes no formal link between the constraints of a language and UG. Less marked constraints will occur in many if not all, languages; more marked ones will occur in fewer languages. Some constraints are so marked as to be peculiar to a single language. Indeed, the lexical entries must be regarded as constraints specific to one particular language. Moreover, although DP depends crucially on underspecification and redundancy rules (constructive processes), it is difficult to reconcile DP with any extant theory of underspecification: In principle, one lexical entry...
may specify particular values for some features, while the opposing values are specified in other lexical entries.

According to OT, phonology consists of a function called GEN, which produces for each input form a set of potential surface candidates (McCarthy and Prince 1993:5), and a universal bank of constraints that is prioritized, from most to least influential, on a language-specific basis. OT's general conception of the grammar is shown in (4).

(4) The Organization of the Grammar in OT:
   Input (underlying forms)
       ↓
   GENeration of candidates
       ↓
   EVALuation of candidates by the ranked constraints

The candidate set that GEN produces is referred to as "potential", because the constraints of the language (EVAL) determine which member of the set will be selected as the output. The constraints decide whether each candidate is well-formed or not, though they often have conflicting requirements and, therefore, often disagree among themselves over what it means to be well-formed (Prince and Smolensky 1993:5). Every case of constraint conflict (to be discussed subsequently) is a case where two constraints have incompatible ideas about which candidate is well-formed (see also McCarthy, this volume).

GEN's production of a candidate set is governed by three principles, which are taken from McCarthy and Prince (1993:20), and shown in (5).

(5) Principles governing the operation of GEN:
   a. Freedom of Analysis: Any amount of structure may be posited.
   b. Containment: No element may be literally removed from the input form; the input is thus contained in every candidate.
   c. Consistency of Exponence: No changes in the exponence of a phonologically-specified morpheme are permitted.

We will be concerned here with (5a) and (5b), which clearly interact. Freedom of Analysis permits GEN to posit any amount of structure or features in its production of a candidate set, provided that each candidate contains the input (Containment).\footnote{These two principles are also used by Prince and Smolensky while (5c), Consistency of Exponence, is a contribution of McCarthy and Prince (1993).}

TCRS and OT both follow Stanley (1967) in assuming negative, positive and implicational constraints, though OT has not addressed the issue directly. For example, Itô and Mester's (this volume) Coda Place Condition, which prohibits the Japanese coda from having a specified place of articulation, and the Mora Sonority Threshold, which prohibits segments lower than nasals on the sonority scale from being linked to a mora in Japanese, are both negative constraints. Itô and Mester's analysis also involves positive constraints, notably the Segment Licensing Condition, which states that all segments must be licensed, and the Segment Head Requirement which, as its name suggests, rules that all root nodes must have a head. There is a third, implicational kind of constraint assumed in OT. An example given by Prince and Smolensky (1993:48) is $\text{FtBin}$ ($\text{Ft} \equiv \text{foot}$ and $\text{Bin} \equiv \text{binary}$), which says that if there is a foot in the representation, then it must be binary. OT claims that "languages differ only (or principally) in constraint ranking [their emphasis], not in the formulation of constraints" (McCarthy and Prince 1993:34). In other words, OT's bank of constraints is universal, and so is their formulation; only constraint ranking is language-specific. If this extends to constraint formulation in positive, negative, or implicational terms, then distinguishing among the three types of constraints may be particularly important for OT, given that positive and negative constraint formulations make different predictions (LaCharité, this volume).

TCRS has been very explicit about what a constraint is and what its characteristics are. For TCRS, constraints delimit universal or language-specific generalizations that account for the inventory, distribution and combinations of phonological elements and structures in a language, as well as phonological alternations. Thus, like DP, TCRS considers that languages differ from one another by virtue of their constraint repertoires. However, in contrast to DP, language-specific constraints in TCRS have a clear link with UG, which is claimed to consist of principles (universal constraints) and parameters for phonological content and structure (Paradis 1988b; see also Kaye et al 1985 for a similar view). Negative settings for parameters constitute negative language-specific constraints. For example, US provides the possibility of nasalized vowels. English is among the many languages that say no to this option, thus it has a language-specific constraint against nasalized vowels. French is among those languages that have nasalized vowels, so it lacks this negative constraint. Still, only some vowels in French can be nasalized ($\tilde{a}, \tilde{e}, \tilde{o}, \tilde{u}$); French says no to the option of having tense nasal vowels, $*\tilde{a}, *\tilde{e}, *\tilde{o}, *\tilde{u}$, $*\tilde{e}, *\tilde{i}, *\tilde{o}$ (though their oral counterpart is perfectly well-formed), resulting in a language-specific constraint (Paradis 1993a). This parameter-setting approach to constraints is illustrated in (6).
neither OT nor TCRS can afford to ignore: given that it is possible to
formalize constraints in a way that avoids constraint conflict, then
there must be justification for not doing so. However, one must always bear in
mind the cost that DP pays for avoiding constraint conflict. Well-formedness
in DP means satisfying all the constraints, so precision in the formulation
of constraints must be allowed to mean great specificity. Not only does this
open the door to ad hoc, idiosyncratic constraints with absolutely no link to
UG, but DP still has to handle constraint conflicts (which is done through
the Elsewhere Condition), as we have just mentioned and as we will see in
more detail in Section 2.3.

In OT, conflict arises when two constraints make competing demands,
i.e., would each select a different output from GEN’s candidate set. For ex-
ample, McCarthy (this volume) describes two conflicting constraints, Final
C and Coda-Cond, that have a bearing on the behaviour of Boston r, an
enepithetic segment which does not always occur where, according to pre-
vious accounts, it is expected. McCarthy accounts for this unexpected be-
haviour in terms of constraint conflicts. Final-C demands that a coda be
supplied for a syllable ending in a short vowel, a role often served in this
dialect by r intrusion (e.g., spa → [spar]). However, according to the Coda
Cond, r can only be in a coda if it is simultaneously linked to a following
onset (i.e., ambisyllabic), which is feasible only when this onset belongs to
the same phonological word as the coda r. Thus, when a short vowel occurs
in the last syllable of a phonological word, Final-C and Coda-Cond conflict.
The candidate that has r in the coda meets the requirements of Final-C,
but must ignore, at the end of a phonological word, the ambisyllability re-
quirement of r insertion; the candidate that lacks r meets the requirements
of Coda-Cond, but leaves the short vowel without a coda, violating Final-C.
It is not entirely clear, however, why in some cases only problematic can-
didates are generated, when GEN is relatively unconstrained in its produc-
tion of candidates, and one can easily imagine candidates which would satisfy
both constraints. In other words, why candidates like spat or spar, which
would be well-formed from the point of view of both constraints, are not
produced by GEN is not clear to us.\(^8\) GEN’s production of candidates is
further discussed in Section 2.5.

The notion of constraint conflict is broader in TCRS than in either DP or
OT. On the one hand, TCRS shares with DP the notion of constraint conflict
where violation is avoided by prioritizing the constraints (Section 2.3). Mor-

\(^8\) McCarthy accounts for r insertion with a morphologized rule that lies out-
side GEN but nonetheless contributes candidates to GEN for evaluation by the
constraints. Our concern is that the rule feeds into GEN an optimal candidate. If
a rule such as r insertion operates this way, why not other rules, for example, a
rule inserting t or p?
specifically, a constraint conflict in TCRS can be created by an element or a structure that is problematic from the point of view of two or more different constraints. For example, the glide /q/ and the fricative /v/ in French [kryv] cuivre 'copper' violate several constraints at the same time when the word is borrowed by Fula, a language lacking /q/ and /v/ and prohibiting branching onsets/codas and rising diphthongs. However, since the constraints, the syllabic and segmental ones, simply have simultaneous (non-divergent) demands on /q/ or /v/ and since, as in DP, there is a general mechanism — in TCRS it is the Phonological Level Hierarchy, discussed in Section 2.3 — for granting priority to one type of constraint (the syllabic one) over the other (the segmental one), there is no resulting violation. Therefore, this type of conflict is of no concern in comparing TCRS and OT, the latter being concerned only with constraint conflicts leading to constraint violations.

However, constraint conflict in the view of TCRS also includes a second, violation-inducing "damned if you do, damned if you don't" situation that arises when applying a repair violates one constraint but blocking (the application of a repair) violates another. Consider the case of floating nasals in French. The nasal in (7a) is problematic because it is unlicensed, i.e., unattached to a timing unit. The most economical way to repair the violation is to spread [+nasal] to the preceding vowel, which has the effect of indirectly licensing the root node of the nasal consonant. This propagation is possible since French allows [-consonantal] and [+nasal] to combine (Paradis 1990b; Paradis and El Fenne 1993). However, if the preceding vowel is tense, as shown in (7b), a violation of the constraint against tense nasal vowels, given in (6), results. We are faced with a dilemma (whose solution is presented in Section 2.3): leaving the root node unlinked violates the licensing requirement; linking the [+nasal] feature to the vowel violates (6).

(7) a. Violation of licensing:  
   \[
   \begin{array}{c}
   V \\
   X \\
   \bullet \quad \text{Root Node} \\
   \quad \quad \quad [+\text{nasal}] \\
   \end{array}
   \]

   b. Violation of * [+tense] [+nasal]:  
   \[
   \begin{array}{c}
   V \\
   X \\
   \bullet \quad \text{Root Node} \\
   \quad \quad \quad [+\text{tense}] \\
   \quad \quad \quad [+\text{nasal}] \\
   \end{array}
   \]

9 Fula adaptation yields /kiri/, where problematic French branching syllabic constituents are simplified, which has also the effect of solving the segmental problems, i.e., the presence of the /q/ and /v/.

10 Having [+nasal] dominated by a Spontaneous Voice Node (e.g., Rice and Avery 1991) and spreading the Node would yield the same results. Though details of the analysis would change, it would readily accommodate to TCRS.

Despite the claim by McCarthy (this volume) that constraint conflict in the two frameworks is very different, it is not difficult to see the connection between this TCRS view of constraint conflict and that of OT. Whether we phrase (7) in OT's or TCRS's terms, the situation is the same: meeting the requirements of the licensing condition means violating the constraint against tense nasal vowels. Alternatively, abiding by the constraint against tense nasal vowels means violating the licensing requirements. Paradis (this volume) provides another example from Gere in which vowel raising — motivated by a universal constraint against diphthongs involving two mid vowels (see Paradis 1990a:64)— may result in a round vowel which, when it follows a rounded (i.e., labialized) consonant, violates the OCP constraint *[-round][+round]. Again, one might say that meeting the requirements of the diphthong constraint means violating the OCP, and vice versa.

2.3. How are Constraint Conflicts Solved?

Even though DP disavows the idea of constraint conflict altogether, the fact remains that all three theories need a mechanism, if not for settling a constraint conflict, then for avoiding it. The mechanism used by each theory is presented in Table 3.

<table>
<thead>
<tr>
<th>Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Declarative Phonology</strong></td>
</tr>
<tr>
<td>• Averted by the Elsewhere Condition</td>
</tr>
</tbody>
</table>

DP uses the Elsewhere Condition to avert a constraint conflict: a more specific constraint takes precedence over a more general one, specificity being defined by inclusion. If α is included in β, it is a special case of β and thus takes precedence. In his DP analysis of Berber syllabification, Scobbie (this volume) invokes two constraints which illustrate this point. Parse ensures that each feature bundle is syllabified as an onset, nucleus or coda; Nexus states that the segment α is dominated by a nucleus. It is thus more specific than Parse and would prevent α from being syllabified as anything other than a nucleus.

OT rejects the idea that a specific rule must always dominate a more general one (see the OT analysis of Lardill offred by Prince and Smolensky 1993:81). OT uses language-specific constraint ranking and the principle that violation is minimal (McCarthy and Prince 1993:1). Though the bank of constraints is universal, a language can prioritize them, giving one constraint priority over another — which allows a language to give
priority to the general, rather than the more specific constraint. When two constraints conflict, the language fulfills the demands of the (language-specified) higher ranked constraint at the expense of ignoring the lower ranked one. This is one interpretation of minimal violation. OT takes what may be characterized as a “lesser of two evils” approach to solving constraint conflicts. In the particular dialect of English discussed by McCarthy (this volume), the constraint Coda-Cond has priority over Final-C; so given that GEN has apparently been unable to generate a candidate that meets the demands of both constraints in the particular situation, the language is satisfied with the form that meets the demands of the higher ranked Coda-Cond, thus accounting for the fact that short vowels may surface without a coda at the end of a phonological word.

TCRS also takes what has just been described as the lesser of two evils approach to solving constraint conflict, though what it means for a constraint to be “lesser” is different in TCRS than in OT. OT defines lesser in terms of a language-specific ranking, whereas TCRS, like DP, invokes an independent and universal mechanism in constraint conflict situations.11 TCRS relies on the universal Phonological Level Hierarchy (PLH), which is simply a reflection of the independently-required hierarchy of phonological organization: metrical level > syllable > skeleton > root node > class node > terminal feature (Paradis 1988b:7).12 When a given structure violates two constraints at once, the needs of the constraint bearing on the higher level in the PLH have priority over those on lower levels. In the French nasal situation referred to in Section 2.2, the licensing needs of an organizing node have priority over the lesser constraint against a combination of features. This accounts for the fact that a tense vowel followed by a nasal alternates with the lax nasalized vowel in French (e.g., brun [bʁɛ] *[bʁɛ] ‘brown’ masculine ~ brune [bʁyn] feminine). More precisely, the feature [+tense] is dissociated from the vowel by a repair strategy after [+nasal] has spread.

There is an important point concerning constraint conflict that is relevant to all three theories. No one wants ordering of constraints, if it can be avoided. Constraint conflict is the only situation to compel ordering, something explicitly noted by OT (McCarthy and Prince 1993:31). Except in cases of conflict, constraints need not be, indeed justifiably cannot be, assumed to be ordered. A major difference among DP, OT and TCRS lies in the mechanism invoked to order constraints in situations of conflict (potential or actual). DP and TCRS rely on (different) intrinsic mechanisms, whereas OT appeals to an extrinsic ordering, which is, a priori, theoretically less desirable. Of course, the mechanisms invoked must first and foremost account for the facts. Further research will hopefully indicate which mechanism is empirically the most efficient.

2.4. Can Constraints be Violated?
As can be observed in Table 4, the question of the violability of constraints constitutes one of the main distinctions between DP, on the one hand, and OT and TCRS, on the other.

<table>
<thead>
<tr>
<th>Declarative Phonology</th>
<th>Optimality Theory</th>
<th>TCRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

In DP, all constraints, including the lexical entry itself, are partial descriptions of the surface form. Except insofar as a more specific and a more general constraint apply disjunctively, all constraints are surface-true, i.e., inviolable. In contrast, constraints are considered violable by both OT and TCRS.

Because this issue is more complex, because the positions of OT and TCRS are so close on important points concerning constraint violation, and because the differences in their positions are sometimes subtle, the two theories will be discussed in tandem, rather than sequentially. Despite possible differences about what constitutes a constraint conflict, both OT and TCRS claim that constraint conflict is a source of constraint violation. In OT, when two language-specifically ranked constraints have incompatible requirements, the lower ranked constraint must be violated so that the requirements of the higher ranked one can be met. In TCRS, repairing a violation of (what is from the point of view of the PLH) a higher ranked constraint violates a lower ranked one (which is reminiscent of constraints having incompatible requirements). TCRS explicitly claims two other internal sources of constraint violation, underlying ill-formedness (Paradis, this volume; and Paradis and Prunet, this volume) and morphological and syntactic operations (LaCharité, this volume). There are also external sources of constraint violations, such as loanwords (Paradis et al. 1993).

Concerning underlying ill-formedness, OT’s position is very close to that of TCRS, which states that constraints are active in the lexical and post-lexical component but not in the DICT (dictionary), i.e., the first input list.

---

11 In DP’s case, as we have discussed, the conflict is averted, rather than solved but the fact remains that this is accomplished by a universal and independent mechanism, the Elsewhere Condition.

12 The hierarchy provided by LaCharité (this volume) and Paradis (this volume) (i.e., metrical > syllable > timing slot > segment > non-terminal feature > terminal feature) represents a more recent theoretical development of TCRS and departs slightly from the hierarchy given here and in Paradis (1990a) and Béland, Paradis, and Bois (this volume).
INTRODUCTION

TCRS framework, LaCharité (this volume) provides an example of a morphological violation: a constraint against [+stressed] and Labial in Setswana is directly violated when passivization links [+round], a Labial dependent to a strident consonant (e.g., disa ‘herd’ → *dishwa → *difwa). OT cannot allow such morphological operations to violate constraints directly, as in TCRS. However, because in OT output forms are selected, not derived, morphological operations can be respectful of phonological constraints and still lead to phonological constraint violation, as follows. Output forms including morphological complexes, are selected from a field of candidates produced by GEN; these candidates include both more and less optimal forms. Thus, OT might say that in the Setswana example, both *dishwa and difwa are among the candidates, but that the constraint against labialized stridents leads to difwa being selected as the optimal output. Since difwa meets the requirements of the constraint at issue in this case, there is no morphologically-created constraint violation. However, a fundamental claim of OT is that in many cases there is no candidate, morphologically-derivable or derived, that can satisfy all the constraints. In other words, there are morphologically-derived candidates which do not meet the requirements of all constraints. In all such cases, a candidate will be selected according to the principle of minimal violation (i.e., the violation of a lower-ranked constraint is preferred to the violation of a higher-ranked one). In the sense that morphological operations can result in surface forms which do not satisfy all constraints, morphological operations must be regarded as potential sources, if indirect, of constraint violations.

2.5. Are Constraint Violations Repaired?

As indicated in Table 5, the question of repair is irrelevant to DP, which rejects the notion of constraint violation, but it constitutes one of the main distinctions between OT and TCRS.

<table>
<thead>
<tr>
<th>Declarative Phonology</th>
<th>Optimality Theory</th>
<th>TCRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

As far as OT is concerned, constraint violations persist. The best possible candidate is always selected, but an output, to be considered well-formed, need not meet the demands of all the constraints. For instance, (/he) shoulda eaten [Juda ijan] ("Juda r ijan) is considered well-formed in the Massachusetts dialect, despite the fact that it violates the constraint Final-C (McCarthy, this volume).
TCRS claims that a well-formed representation must conform to all the active constraints of the language. If constraints conflict, their needs are prioritized by the PLH, which determines only the order in which constraints are dealt with. Ultimately, the needs of all constraints are met through the mechanism of repair, which is given in (10).

(10) Repair:
A universal context-free phonological operation that inserts or deletes content or structure to make a phonological unit or structure conform to a constraint.

Repair is universal in terms of its principles of operation, though the particular material upon which it operates is attributable to a constraint, which may be language-specific (i.e., a negative parameter setting). One universal principle governing repair is the Preservation Principle, given in (11) (Paradis, this volume; Paradis et al 1993).

(11) Preservation Principle:
Preserve as much of the input as possible, according to the constraints of the language.

The Preservation Principle is a recent addition to TCRS, though dissenting claims of the theory have always conspired to suggest its existence. It explains why, in a constraint conflict, priority is given to the constraint bearing on the higher level of phonological structure, according to the PLH. The higher the level in the phonological hierarchy, the more phonological information is at stake. The Preservation Principle also explains why blocking a violation-inducing operation has precedence over going ahead with it, even though a violation can subsequently be repaired (Paradis 1988a:75). According to TCRS, a violation cannot persist; it must be repaired. But a repair strategy, by definition, adds to or subtracts information from the phonological form to which it applies, which results inevitably in a transformation of the input. Therefore, avoiding violations (and their subsequent repairing) is the preferred option (it preserves the input most effectively), even though repair is available to mend them. The Preservation Principle further accounts for the fact that insertion/linking very often has priority over deletion/de-linking (Paradis and Prunet 1988:3; Paradis, this volume). Insertion/linking has the effect of adding information to an input, whereas deletion/de-linking has the effect of subtracting (destroying) phonological information from an input (see Paradis et al 1993 for details on this particular point). Finally, the Preservation Principle provides the rationale for the Minimality Principle given in (12) (see in this volume Béland et al LaCharité; Paradis; and Paradis and Prunet).

(12) Minimality Principle:
A repair must apply at the lowest phonological level to which the violated constraint it preserves refers.

The Minimality Principle may be thought of as operationally interpreting the Preservation Principle by defining what it means to preserve the input as far as possible (or, conversely, to lose as little as possible). If phonological material is deleted, it is at the lowest level of a problematic structure, according to the PLH; if phonological material is inserted, it is the least amount of phonological information possible, as determined by the PLH. A priori, adding a feature to or subtracting it from an input alters a morpheme's phonological makeup less drastically than adding or subtracting a segment, syllable, foot, etc.

Though OT does not have the notion of repair, the principles that govern the function of GEN, notably Containment, are somewhat reminiscent of the Preservation Principle that governs repair. There is, however, at least one crucial and questionable difference: in TCRS, repair is a response to a constraint violation; if an input is found wanting with respect to the constraints, it undergoes repair according to the Preservation Principle and the Minimality Principle. In OT, the changes made by GEN are not made because the input violates or anticipates a constraint, even though the sole function of GEN seems to be to provide a candidate set from which the constraints may select an optimal form (i.e., one that satisfies the constraint hierarchy). In any candidate set, all but one of the generated forms are gratuitous. Herein lies a problem: changes in GEN are not made in response to constraint requirements, i.e., they do not constitute repairs (the generation of an optimal form is fortuitous with respect to the constraints). Therefore, GEN must be allowed to seriously overgenerate by producing many non-optimal forms. However, this overgeneration of GEN, which is intended by the proponents of OT, is an extremely costly and unexplanatory device. It departs from previously accepted scientific criteria (economy, power of prediction, links between facts, etc.) which normally serve to evaluate the adequacy of a phonological analysis (however, see Prince and Smolensky 1993:197–198 for a brief discussion).

---

14 Even though the Minimality Principle has been formalized for the first time in this volume, the notion of minimality has always been present in TCRS. See, for instance, Paradis and Prunet (1988:6), where a repair strategy is defined as a "minimal phonological process" and Paradis (1990:56), where it is said that "repair strategies apply minimally, according to constraint focuses".
In sum, although McCarthy and Prince (1993:20) acknowledge that "it is essential to establish what a candidate set actually consists of: to define the function GEN", the internal workings of GEN are largely unexplored and seem subject to two conflicting criteria. On the one hand, the constraints on GEN cannot be too strict, for without a rich candidate set to choose from, only extremely good fortune can account for the actual output, the optimal candidate. On the other hand, the same is true if GEN's principles are abandoned altogether and the candidate set is allowed to become too rich. After all, why must we ever settle for an output that violates a constraint when an infinite field of candidates would make available forms that conform to all the constraints? Even if we accept the principles which constrain the generation of a candidate set, we must ask what independent value lies in generating a multitude of candidates, only one of which will be an output form. Clearly, consideration needs to be given to the price of avoiding repair.

2.6. How are Surface Exceptions Accounted For?

When all is said and done, no one can maintain that constraints are entirely without exception. DP uses the Elsewhere Condition to explain surface exceptions. Any apparent violation of a constraint leads us to assume that a more specific one, which renders the more general constraint inapplicable, is at play in the particular instance. The advantage of this view is that both constraint prioritization and surface exceptions are accounted for by a single, independently-motivated mechanism, the Elsewhere Condition. The disadvantage has already been pointed out (Section 2.1): constraints in DP must be allowed to be very specific and without a principled link to UG. It is also unclear how such specificity can be falsified.

To explain surface exceptions to constraints, OT relies on the idea that a higher ranked constraint has priority over a lower ranked one, making the needs of the lower ranked constraint irrelevant from the point of view of optimality. However, ranking of constraints is not just language-specific, but level-specific. There is, in OT, a second tool—lexical levels—to explain constraint behaviour. Though there is little discussion of lexical level-ordering these days, the notion of strata persists and is common to both OT and TCRS (Paradis, this volume), according to which a given constraint may be influential at one stratum and not others. Because in OT the bank of constraints is claimed to be universal and therefore common to all levels, lack of adherence to a particular constraint is explained in terms of re prioritization, which can occur at each level. A given constraint may be highly ranked at one level, meaning that it will usually, if not always, have its needs met at that level, while being low ranked at another, leading to its more frequent violation. Even though OT provides generally formulated and universal constraints, as does TCRS, language-specific ranking and reranking seems dangerously ad hoc and relatively hard to falsify.

TCRS appeals to the notion of a constraint domain to explain surface exceptions to constraints. This means that a constraint can apply at certain lexical strata and not others, including the surface (Paradis 1988a, 1988b: Paradis and Prunet 1988:1). The domain of a constraint must include contiguous lexical strata, so that a constraint cannot be arbitrarily turned off and on (see the Stratum Domain Hypothesis in Mohanan 1986:47). Thus both OT and TCRS allow for intermediate forms whose well-formedness is referenced to a particular phonological level.

Table 6 summarizes the (universal and language-specific) tools used by the three constraint-based theories discussed in this volume. This tableau may give the impression that DP, the framework with the fewest tools, is the most efficient. However, economy (of tools) is achieved at the cost of an important loss of predictive power. One must not forget, as already seen in Section 2.1, that constraints in DP are always language-specific, and can be extremely ad hoc, and that DP resorts to an undetermined theory of underspecification, which is extremely powerful. DP also depends heavily on phonetics to account for alternations that cannot be handled by redundancy rules/constraints (see Russell 1993). Moreover, the criticisms addressed to Kaye et al.'s (1985) use of constraints can also be directed at DP, since DP's constraints, like that of Kaye et al, both specify the structure and accomplish the change (here, the insertion of information).

---

11This is true also of Harmonic Phonology, following Goldsmith (1991, 1993, etc.)
3. Conclusion
In Section 1, we have provided a retrospective on the emergence of constraints within generative phonology because, regardless of theoretical perspective, we need to be mindful of crucial issues facing us as we try to replace arbitrary rules with constraints and to establish the rightful place of phonological constraints in the grammar. For theorists in any constraint-based framework, this means considering the views of those who have dealt with the issues previously, who have not been bound by the same pragmatic considerations as we ourselves are. In Section 2, we have endeavored to master the unique vocabulary of each of the three constraint-based theories presented in this volume—DP, OT and TCNS—to provide an overview of what we see as essential points of consensus and contrast among them. If our interpretation is correct, there are as many important parallels as there are differences among the three theories, though OT and TCNS are perhaps more similar to each other than to DP. We hope that sorting out the differences and the similarities as we have done in this introduction will enable the reader to understand better the papers in this volume and will help theorists within all three frameworks to test the real points of difference, and not be misled by different terminology. In a more general fashion, we hope that this volume will stimulate further research on constraint-based theories.

REFERENCES

Archangeli, Diana

Bird, Steven, John Coleman, Janet Pierrehumbert, and James Scobbie

Chomsky, Noam, and Morris Halle

Clayton, Mary L.

Coleman, John
1993 Declarative Lexical Phonology. Ms.

Dresher, Bezalel Elan

Goldsmith, John

Goldsmith, John, ed.

Goldsmith, John, and Gary Larson

Halle, Morris

Hockett, Charles

Hooper, Joan

Itô, Junko

Kahn, Daniel

Kaye, Jonathan, Jean Lowenstamm, and Jean-Roger Vergnaud

Kenstowicz, Michael, and Charles Kisseberth

Kiparsky, Paul

Kisseberth, Charles

Lehmann, Winfred P.

McCarthy, John
Paradis, Carole

Paradis, Carole, and Fatim El Fenne

Paradis, Carole, Caroline Lebel, and Darlene LoCharité

Paradis, Carole, and Emmanuel Nikiema

Paradis, Carole, and Jean-François Prunet

Postal, Paul

INTRODUCTION

Prince, Alan, and Paul Smolensky

Fulleyblank, Douglas
1986 Tone in Lexical Phonology. Dordrecht: Foris.

Rice, Keren and Peter Avery

Russell, Kevin

Scobie, James


Shibatani, Masayoshi

Singh, Rajendra

Stanley, Richard

Yip, Moira