NATIVE AND LOANWORD PHONOLOGY AS ONE: CONSTRAINTS VS. RULES*

Carole Paradis
Laval University, Quebec, Canada

ABSTRACT
Constraint-based theories have gained increasing recognition over the past four years. This paper aims to show the superiority of one of these theories, the Theory of Constraint and Repair Strategies, over a rule-based approach with respect to loanword behavior. While the latter requires phonology to be split into two sets of language-specific processes — one for loanwords and another one for native words — the former proposes a unique set of universal processes.

1. INTRODUCTION
Constraint-based theories — which are characterized by the rejection of arbitrary "rules" — certainly constitute the liveliest area in current linguistics. Among these theories the best known in phonology are Optimality Theory (Prince & Smolensky 1993), Harmony Phonology (Goldsmith & Larson 1992; Goldsmith 1993), the Theory of Constraints and Repair Strategies (TCRS) (Paradis 1988a,b; Paradis & LaCharité 1993) and Declarative Phonology (Scobbie 1991; Bird et al. 1993). Except perhaps for proponents of the latter, whose primary purpose is the computerization of phonology (cf. Paradis & LaCharité 1993), proponents of constraint-based theories claim that constraint-based accounts are more explanatory than those couched in rule-based theories — the first being that of Chomsky & Halle (1968) — in the sense that the former capture more generalizations, remove a great deal of arbitrariness and redundancy from analyses, and make more predictions on language-specific and universal bases.

This paper defends this view — which is already strongly supported by internal (native) evidence — within the framework of TCRS, and on the grounds of loanword adaptation.

Constraints are often detected when violated because a violation normally yields a deviation from what would otherwise be expected. TCRS identifies morphological operations as the main source of constraint violations. For instance, consider the case of the mid vowel [o] in French, which never occurs word-finally. Its absence can be interpreted in two ways: as an accidental gap or as evidence for a constraint against [o] in word final position. The second option is selected because there is what I will call "dynamic" phonological evidence provided by the morphology of French supporting it. Two pieces of evidence come from the vocalic alternation found in adjectival inflection such as sur [sor] (as oile [oile] 'stupid (masc./fem.)' and in verbal derivation such as complet [kplkom] ("[kplkom] 'plot/complot [kplkom] 'plot'. Note that the existence of such morphemes as chaud [fou] (chau def) 'hot (masc./fem.)' and in verbal derivations such as endos [edos] 'endorsement' (endosse (edosse) 'to endorse' in French — where the vowel [o] is realized in word-final and non-final position — clearly indicates that the prohibition bears on the vowel [o] in word-final position, not the vowel [o] in non-final position. Abbreviation constitutes another source of evidence for this constraint: professionel [profesjonel] (profesjon) 'professional', Carole [karol] (Car0 [kar0]) 'Carole', police [polis] 'popo [popo] (popo) 'police'. In all cases, the underlying vowel [o], which surfaces in non-final position, yields [o] at the end of abbreviations, thus showing that the process is too general, i.e. it occurs in too many different morphological contexts, and too systematic (there is no exception) not to be the result of a phonological constraint against [o] in French.

However, it is common for linguists to be left only with "static" evidence, i.e. the absence of an element or structure x in a given language, to suspect the existence of a constraint. For instance, it can be observed that #CC sequences do not exist in Fula, a West-African language. It is tempting for a linguist to resort to a constraint to express this fact, but one does not know with certainty whether the lack of such a structure is due to an accidental gap in the language, a diachronic constraint or a synchronic one (cf. Paradis & Prunet 1993). Derivation and inflection of native words do not provide any insight here since there is no morphological operation in Fula which would generate such a sequence, i.e. there is no mono-consonantal prefix which would attach to a consonant-initial word, and thus yield a #CC cluster.

This is where borrowings play a crucial role: they often contain elements or structures that are absent from the native vocabulary. Depending on how these foreign elements or structures are treated by the borrowing language — is it accepted or systematically modified (adapted)? — the linguist may know whether the absence of such elements or structures in the studied language is due to a constraint or an accidental gap. For instance, Fula has borrowed extensively from French, a language with branching derivations. Adaptation of French borrowings with such onsets provides dynamic evidence for the existence of a constraint against #CC clusters in Fula since all such French clusters are automatically modified in Fula. They usually yield #CCV sequences, i.e. sequences with a vowel inserted in between the two consonants (e.g. Fr [en]ch [t]racteur [traktor] 'tractor' F (ula) [t]raktor and Fr place [plas] 'place' F [plas] place). From the perspective of TCRS, borrowings constitute an invaluable source of constraint violations, which allow the linguist to observe how a language "reacts" to unfamiliar elements or structures.

Paradoxically, however, the fact that these phonological "reactions" are sometimes restricted to loanwords — for the reasons we have just seen in Fula, i.e. there is sometimes no context in the language from which a constraint violation might stem, and thus no possible "reaction" to violations — has led some linguists to conclude that there were two separate sets of phonological processes, one for loanwords and one for native words. Silverman (1992) is among the ones who maintain this view the most explicitly. Such a position, which stems from a rule-based perspective, is at best useless in a constraint-based perspective (cf. also Yip 1993: 262). I will show that what I call the Two Process-Set Hypothesis, in (1), entails non-desirable effects such as duplicating identical processes in the same language and, above all, missing important links among facts, on language-specific and universal grounds.

(1) Two Process-Set Hypothesis: Loanwords and native words each have their own set of processes (rules).

To this effect, we will examine three constraints (CC#, *CC, *VV) each in a different language (Fula, Kinyarwanda and Moroccan Arabic, respectively), and observe how the processes triggered by these constraints would have to be handled in a rule-based approach. The paper will be organized as follows. Section 2 presents my assumptions regarding borrowings (2.1 and 2.3), and the relevant tenets of TCRS (2.2). Section 3 addresses the three constraints mentioned above, while section 4 offers a brief conclusion.

2. ASSUMPTIONS

2.1 Borrowings

Two opposite views are debated in loanword studies: the "phonetic approximation stance" (e.g. Haugen 1950 and Silverman 1992), where a borrowed word is analyzed as a non-linguistic acoustic signal, and the "phonological stance" (e.g. Hyman 1970 and Prunet 1990) where a borrowed word is instantaneously assigned a mental representation in the recipient language (L1). Strong arguments based on sociolinguistic, psycholinguistic and phonological studies have been recently brought forward by Paradis et al. (1995a,b) in favor of the phonological stance. For instance, sociolinguistic studies (e.g. Haugen 1950 and Poplack et al. 1988) clearly indicate that borrowings are introduced by bilinguals (not monolinguals), who have ac-

*I am deeply indebted to Heather Goad and Jean-François Prunet for precious comments. Thanks also to Yan Rose for his help with the editing of the paper. The author remains solely responsible for any remaining errors or omissions. I acknowledge SSHRC grant # 410-94-1256 and FCAR grant # 95-ER-2305.
cess to the phonology of the source language (L2). Loanwords are introduced by bilinguals through what sociolinguists call "code-switches", "nonces" and "idioms". Sociolinguistic studies also clearly show that phonological patterns of adaptation are imposed by bilinguals; they are community-wide, especially in mid and high community bilingualism stages. This indicates that borrowing integrators and adapters have access to word representations in L2.1 Otherwise adaptations could not display the strong consistency observed by Haugen (1950) in the mid and high community bilingualism stages and by us in our own corpora of loanwords (cf. Paradis et al. 1993, 1995a, b for a thorough argumentation in favor of the phonological stance).

2.2 Framework: TCRS
In TCRS, a language's phonology consists of both universal and non-universal constraints which, when violated, trigger the application of a repair strategy (e.g. *a# → o in section 1), defined in (2).

(2) Repair strategy: A universal, context-free phonological operation that is triggered by the violation of a phonological constraint, and which inserts or deletes content or structure to ensure conformity to the violated constraint.

As mentioned in section 1, TCRS claims that constraint violations originate mainly from morphological operations (e.g. the constraint *a# discussed in section 1, which is violated because of an abbreviation operation, etc.). Other internal sources include constraint conflicts and underlying ill-formedness (Paradis 1988a, b). Loanwords (Paradis et al. 1993) and paraphasias (Béland et al. 1993) constitute external sources. However, while violated constraints press for repair, the Preservation Principle, (3), protects the input, i.e. resists segmental loss.

(3) Preservation Principle: Segmental information is maximally preserved within the limits of the Threshold Principle ((4)).

I maintain that the Preservation Principle is responsible for the low rate of segment deletion observed in the four corpora of loanwords that we have built (4,031 borrowings from French into Kinyarwanda, Moroccan Arabic and Fula, and English into Quebec French), which contain altogether 12,630 malformations. The Preservation Principle works in the following way. Repair is accomplished by the insertion or deletion of content (e.g. features, timing units, etc.) or structure (links between features, various levels of structure, etc.). At its most basic, repair by insertion occurs when a constraint violation is due to a lack of content or structure whereas deletion applies when a constraint is offended by an excess of content or structure. Whether a problem is due to a lack of something or an excess of something is often a matter of perspective. For example, in a language with a constraint against consonant clusters (CC) such as Kinyarwanda, a CC (loan) input can be regarded as an excess of consonants, leading to deletion (of a consonant), or as the lack of a vowel, leading to insertion (of a vowel). All else being equal, the Preservation Principle, which resists the loss of phonological information, favors viewing a problematic structure as a lack of content or structure, giving preference to insertion over deletion.

TCRS nevertheless posits limits to preservation, i.e. to the price languages are ready to pay to conserve segmental information. This is expressed by the Threshold Principle in (4).

(4) Threshold Principle:
(a) All languages have a tolerance threshold to segment preservation.
(b) This threshold is the same for all languages: two steps (or two repairs) within a given constraint domain.2

The Threshold Principle stipulates that a problematic segment requiring more than two steps to be adapted within a constraint domain — a constraint domain being simply the scope of a constraint violation — is not protected by the Preservation Principle.

Repair, be it by deletion or insertion, must nevertheless apply economically. Economy is expressed first and foremost by the Minimality Principle in (5).

(5) Minimality Principle:
(a) A repair strategy must apply at the lowest phonological level to which the violated constraint refers.
(b) Repair must involve as few strategies (steps) as possible.

The "lowest phonological level" referred to in (5a) is determined by the Phonological Level Hierarchy (PLH), in (6), which simply reflects the phonological organization required independently of TCRS.

(6) Phonological Level Hierarchy:
Metrical level > syllabic level > skeletal level > root node > feature with a dependent > feature without a dependent.

The Preservation Principle in (3) is served by (5a) which minimizes alteration of the input, for example disallowing the loss of a syllable, if the loss of a segment will correct the problem. In other words, it ensures that a constraint violation is solved with as little loss of phonological information as possible. (5b), for its part, requires that, given more than one possible way of repairing an ill-formed structure, priority be given to the repair involving the fewest steps.

TCRS maintains that the phonological structure of a language results from principles (universal constraints) and parameter settings. Principles describe what is common to all languages, whereas parameter settings handle differences (contrasts) among languages (cf. Chomsky 1986). In TCRS, parameters are marked options offered by Universal Grammar. The default reply for a language is to say "no" to such an option, which results in the rejection of a given type of complexity, and thus a negative constraint in the language in question. In this perspective, the segmental inventory of a language is viewed as the direct result of positive and negative language-specific answers (settings) to segmental options offered by Universal Grammar (parameters). In the case of borrowings, one can thus hypothesize that the reason why French coupon [ku'pu'] is realized as [ku'pa'n] in English, i.e. with a (partly) denasalized vowel followed by a nasal consonant is because English says "no" to the following parameter.

(7) Phonemic nasal vowels?
French: yes
English: no (default = constraint)

The negative parameter setting in (7) explains why nasal vowels introduced into English through loanwords are adapted. In the view of TCRS, the recasting of a into a VN shape is not the result of a rule specific to loanwords — as would be the case with the Two Process-Set Hypothesis — but of a constraint active throughout the phonology of English, whose only source of violation is loanwords. This position, that I call the One Process-Set Hypothesis, is formalized in (8).

(8) One Process-Set Hypothesis:
Phonology has access to a single set of two universal processes: insert x and delete x. These processes are repair strategies, whose sole purpose is to yield constraint satisfaction. If there is no constraint violation, they do not apply.

2.3 Core and Periphery
The One Process-Set Hypothesis does not imply, however, that the phonological behavior of loanwords and native words is identical in all respects. If we consider again the case of nasal vowels introduced into English, we realize that while nasal vowels are totally absent from native English words, they are sometimes tolerated in borrowings (e.g. Fr. entrée [entré] → English [entri] or [dtrri]). In a study of loanword adaptation, it is crucial to distinguish between "prohibited" segments, i.e. segments that are systematically and immediately adapted or
eliminated as soon as they are introduced into a language (e.g. the French front round vowels y and ι in English), and "tolerated" segments, which are sometimes adapted and sometimes not (such as the French nasal vowels in English) at least in some speech registers. The latter are called "import" in the literature (cf., e.g., Hauge 1950). To account for the distinction between prohibited and tolerated segments, the TCRS loanword model proposed by Paradis et al. (1995b) views the phonology of a language as being organized into domains. Essentially, a distinction is drawn between the "core" and the "periphery." The core contains all of a language's constraints; by and large, the core defines the phonology of a language and governs its vocabulary. However, not all items in a language are part of the core; some, such as interjections, onomatopoeia, proper names and learned words, along with (partly) unassimilated borrowings, may lie in the periphery, temporarily or even indefinitely. The periphery contains a subset of a language's constraints, which means that items in the periphery are not subject to all the constraints that govern the core. That is to say, the parameter settings for some Universal Grammar options may be set to "yes" rather than "no" in the periphery or some subdomains of the periphery, which effectively deactivates those particular constraints, and accounts for imports (unassimilated foreign sounds). The distinction between core and periphery is not particular to TCRS. It was suggested by Chomsky (1986:147), and further developed by Itô & Mester (1993). However, the core and the periphery are not definite in nature. The periphery is not governed by "new" constraints, i.e., constraints different from those of the core. It contains only "fewer" constraints than the core. In this view, a "borrowing" can be defined as in (9).

(9) Borrowing: An individual word, or compound functioning as a single word, from L2 that a) phonologically conforms to (at least) the outermost peripheral phonological constraints of L1, b) has a mental representation in L1, and c) is incorporated into the discourse of L1 (cf. Paradis et al. 1995a,b for more details).

3. CONSTRAINTS VS. RULES
3.1 Language-Specific Issues
Words in Fula never surface with a final CC cluster or an internal CCC one. We know that this is due to a constraint against branching codas because when such a cluster is present unerringly or arises in the course of a morphological derivation, it is immediately split into different syllables as in (10).

(10) Native words in Fula
fsoof-re → fsoof-e-re 'breath'
talk-ru → talk-u-ru 'amulets'
lacc-ri → lacc-i-ri 'couscous'
The constraint is formally expressed by the negative parameter setting in (11) (recall from section 1 that Fula does not allow branching onsets either).

(11) Parameter:
Branching non-nuclear constituents?
French: yes Fula: no (constraint)
As shown in (12), the constraint also applies to loanwords since CC# clusters in those words undergo vowel insertion too.

(12) French Loanwords in Fula
Fr. carde [karde] → F. karda 'card (comb)'
Fr. force [fɔrs] → F. fɔrs Fula 'force'
Fr. gendarme [ʒadarm] → F. san'darma 'gendarme'
With the Two Process-Set Hypothesis, one would have to posit two separate rules, as in (13), even though both rules would be identical.

(13) Two Process-Set Hypothesis:
a) native words: Φ → V/CC#C
b) loanwords: Φ → V/CC#C
This reduplication of identical rules is seriously flawed in two ways. First, it complicates the grammar. Second, it does not formally capture the fact that both rules are actually the same process (vowel insertion) which is triggered by the same context ([#, C]) in loanwords as in native words. These disadvantages are eliminated with the One Process-Set Hypothesis. As shown in (14), the data in (10) and (12) necessitate only one context-free universal process, i.e. insertion of x.

(14) One Process-Set Hypothesis:
native and borrowed words: Φ → x
While the Preservation Principle in (3) ensures that insertion will have priority over deletion, the Minimality Principle guarantees that material inserted will pertain to the level to which constraint (11) refers (cf. (5a)). Since (11) refers to the syllabic level, the repair will apply at that level. Insertion of a nucleus is selected because this is the only repair which fully satisfies both principles, the Preservation and Minimality Principles. The empty nucleus is subsequently filled by vowel spreading.
The rule-based approach, in which the Two Process-Set Hypothesis is couched, is problematic in other respects. Consider the French borrowings in (15), where vowel insertion occurs in between the two consonants of a CC# cluster, not at the end of it as in (12).

(15) Fr. contre [k5tɛr] → F. kœntœr 'against'
Fr. filtre [filɛr] → F. fîlœr 'filter'
Fr. tableau [tabalo] → F. taabalo 'table'
Not only would the Two Process-Set Hypothesis require the reduplication of the same rule as in (13), the rule-based approach in which it lies, more generally, would require positing a third rule — shown in (16) — to account for the facts in (15).

(16) Φ → V/C_C#
This new rule would be needed because the context of rule (16b) is not identical to that of rule (16). Again, the fact that the trigger is a CC# cluster would be missed. This generalization is straightforwardly captured by constraint (11), however. CC# clusters are prohibited because they would form an illicit branching coda. The insertion locus of the vowel depends entirely on the sonority of the cluster. It is determined by universal markedness, which disfavors [leuʃ] consonant contacts where an onset is more sonorous than the preceding coda, even though such clusters are found in some Fula native words (e.g. faabru 'toad'). In other words, in the absence of opposite morphologically-induced specifications, default settings, provided by Universal Grammar, apply.

From this perspective the phonological behavior of loanwords tells us significantly more about universal default settings than that of native words, which is often morphologized or heavily influenced by diachrony. Once distorting factors such as orthography, analogy, etc. are clearly identified and discarded, one can easily state that loanword phonology is the "emergence of the unmarked" in phonology (cf. McCarthy & Prince 1994 on this notion).

In a rule-based approach, a fourth rule would even have to be posited. As explained above, the Preservation Principle gives precedence to vowel insertion over consonant deletion. However, consonant deletion does occur in a few cases such as those in (17), where v is lost.

(17) Fr. pieuvre [pijuɛʁ] → F. pijœr 'octopus'
Fr. cuivre [kuвиɛʁ] → F. kœri 'copper'
However, as shown in Paradis et al. (1993, 1995a,b), consonant deletion is not random. It is always caused by the presence of an ill-formed segment — here the voiced labial fricative *v*- contained within an unsyllabifiable cluster. Preservation of the two cluster consonants would be too costly in these cases; it would necessitate too many steps (repairs). It would require nuclear insertion and filling as in (10), (12) and (15). But it would also require a third step, i.e. the adaptation of the ill-formed segment *v*- itself (*v* normally yields w in Fula; e.g. Fr. verre [ver] → F. vœr), since it is encompassed within the scope of constraint (11). This would clearly violate the Threshold Principle in (4), which establishes that the limit to segmental preservation is two repairs, within a given constraint domain. Thus not only does TCRS account for the variation in the insertion point of an epenthetic vowel in (12) and (15) without any extra language-specific device, but it also handles straightforwardly the variation in the processes themselves, i.e. insertion of a vowel ((10), (12) and (15)) vs. deletion
of a consonant in (17) (cf. Paradis et al. 1995a,b for more examples and a thorough discussion of these cases). In contrast, a rule-based approach is unable to economically capture this variation, as well as being unfit to perceive the link between the numerous rules it would require to handle the data presented in this section.

3.2 Universal Issues

A rule-based approach would be undesirable on universal grounds also. It would treat the processes observed in the previous section as idiosyncrasies of Fula, despite the fact that restrictions on branching codas are common among languages. Such restrictions are found in Tigrinya and Classical Arabic, for instance. This fact is predicted by TCRS since constraints in TCRS’ view stem from negative parameter-settings. Since parameters are optional offered by Universal Grammar, it is predicted that a number of languages will share the same parameter setting, be it positive or negative. Recall from 2.2 that negative parameter settings are default (unmarked) options; they consist in a language’s refusal of a given type of complexity. Negative settings are thus expected to be relatively frequent.

The same is true of the *CC constraint in Kinyarwanda, a Bantu language, and the *V constraint in Moroccan Arabic, which respectively prohibit codas and empty onsets. Both constraints, which are formalized in (18a) and (18b) respectively, are common across languages. For instance, the former is found in Luganda as well as in most Bantu languages, while the latter is found in Tigrinya, Biblical Hebrew and many other Semitic languages.

(18) a) Parameter: codas?
French, English: yes
Kinyarwanda, Luganda: no
b) Parameter: empty onsets?
French, English: yes
Moroccan Arabic, Tigrinya: no

The constraints in (18) are supported internally, and also externally by the behavior of borrowings like those in (19a) and (19b).

(19) a) French —> Kinyarwanda:
client [kl’al] —> [umu-crija]
citizen [sizor] —> [sito] [sito]
b) French —> Moroccan Arabic:
architect [arbit] —> [laborit]
ascendant [asosser] —> [senser]

In (19a), we can see that French CC sequences are automatically separated by an epenthetic empty nucleus in Kinyarwanda, to which the following vowel spreads. In Moroccan Arabic, a violation of (18b) triggers more diversified repairs, i.e. either insertion of a consonant, as in (19b1), or deletion of the initial vowel as in (19b2). Selection of one repair over another here is conditioned by the length of the output (cf. Paradis et al. 1995b). The longer the output in L1, the more likely vowel deletion is. Nonetheless, both strategies fully preserve (18b) in preventing a vowel from surfacing word-initially. Again, this principle diversity of repairs could not be captured in an explanatory way in a rule-based approach. In such a framework, two completely unrelated rules would have to be posited, thus failing to express the fact that the trigger (*#V) is identical in both cases.

4. CONCLUSION

This paper has attempted to show the superiority of constraints over rules in general. More specifically, TCRS and the traditional rule-based approach of Chomsky & Halle (1968) — which continued to be used under different forms in multilinear phonology and pre-constraint-based frameworks — were compared in their capability to deal with loanwords. The former has proved markedly more economical and explanatory. In particular, it has rendered the Two Process-Set Hypothesis — where loan words and native ones are considered to be each governed by a distinct set of processes — vacuous. On more universal grounds, it was shown that the processes applying to borrowings and native words are non-language-specific idiosyncrasies but the result of the language’s replies to options offered by Universal Grammar, i.e. parameters. The phonological behavior of borrowings, which seems ad hoc in a rule-based view, proves very regular and predictable in TCRS. On the one hand, TCRS provides linguists with a formal framework which handles straightforwardly one of the richest sources of dynamic evidence for constraints: borrowings. On the other hand, the study of borrowings opens a large window on the general functioning of constraints, and ultimately the organization of the language in the human brain, by allowing us to observe how languages react to foreign elements.

REFERENCES