Implication & Impossibility in Grammatical Systems

What it is & How to find it

Alan Prince
October 25, 2006

Abstract. Predictive linguistic theory generates a rich pattern of implications tying together the well- and ill-formedness of linguistic structures. Visible extensionally as restrictions on the kinds of processes and optima that can and cannot co-exist within a grammar, such implications emerge from the constraint ranking patterns that lead to optimality and failure. The *Elementary Ranking Condition* supports a set of practical tools for digging systematic implications out of data as construed by grammatical assumptions. This paper discussses a variety of concrete cases and places them in their analytical context. No background in the relevant notions and techniques is assumed.

Acknowledgments. Thanks to John McCarthy for pointing me toward Anttila's recent work with ERC theory in mind and to Bruce Tesar for an inspiring conversation. I am grateful to Paul Smolensky and Vieri Samek-Lodovici for detailed comments, and to Crystal Akers, Daniel Altshuler, Carlos Fasola, and Michael O'Keefe for helpful remarks on an earlier version of this paper, and to Akin Akinlabi, Birgit Alber, Eric Bakovic, Adrian Brasoveanu, Jane Grimshaw, Ed Keer, Naz Merchant, and Paul de Lacy for valuable discussion of related matters.

Table of Contents

Implication & Impossibility in Grammatical Systems

What it is & How to find it

Alan Prince
Rutgers University
October 25, 2006

0. Overview

Contentful linguistic theories exclude as well as admit. Exclusion can be total, or contingent upon some other state of affairs. The implicational mode is perhaps most familiar from the discourse of ‘markedness’,[1] but as a logical structure and target of explanation it pervades the field. This paper aims to provide some basic tools for exploring such patterns. The key observation developed in Prince & Smolensky 1993/2004 is that relations among forms translate into relations among the ranking conditions that determine the status of those forms. With deepened understanding of the logic of ranking conditions, it has become easier to construe their import and their dependencies.

We pursue a several-pronged attack on the issues involved. The aim is to be sufficiently concrete that the reader who works through the examples will be able to export the methods of analysis demonstrated here. We do not, however, hop from case to case. We wish to set the specific problems addressed within the larger context that gives them significance. We begin by addressing matters so basic as to be thought mere annoyances, showing how they lead to the existence of meaningful structure in grammars (§1-2). We turn then to the analysis of an example recently revived in Anttila 2006, developing the basic tools from scratch along the way (§3-4). We continue by exploring the sense behind the tools and their connection with such familiar notions as harmonic bounding (§5-8). We then resume the analysis of cases (§9), after which we return to the fundamental issue of the relation between formal grammars and the grammatical systems they generate (§10-11).

Check list of abbreviations and notations introduced in the text

OTCIGG: *Optimality Theory: Constraint Interaction in Generative Grammar*

ERC	elementary ranking condition
POC	possible onset condition
CDL	cancellation/domination lemma
WTF	whether to fuse
p,q,x,y	forms p,q,x,y (roman face)
⟨x→y⟩	the candidate with input x, output y.
p q,x,y	candidates *p q,x,y* (italic face), where a candidate is an input-output pair of the form $a_1 \rightarrow a_2$
$x \succ y$	*x* is better than *y*
$[x \succ y]$	the conditions under which *x* is better than *y*, the ERC associated with the assertion $x \succ y$.
α, β	ERCs
⟨1a⟩	candidate 1a
[1ca]	the ERC derived from comparing desired optimum (c) to candidate (a) in candidate set 1.
[1a]	the ERC derived from comparing a desired optimum *known from context* to suboptimum ⟨1a⟩

---

[1] For a recent comprehensive investigation, see de Lacy 2006.

1. Of Numerosity

A constraint set with *n* constraints grants existence to some *n*! constraint hierarchies.[2] The number of distinct *grammatical systems* thereby predicted is typically far, far less than this often grandiose quantity: hence, the emergence of narrow, contentful typologies. *How could this be*?

It proves unexpectedly illuminating to contemplate this question. Two factors are at play, of which the first has drawn the most attention:[3]

- **Within a candidate set**. Severe conditions of symmetry are required to obtain anything like the potential number of *n*! possible optima — conditions that are never even approximated in reality.
- **Across candidate sets.** There are many *incompatibilities* between potential optima. Looking over the entire collection of candidates sets in anything but the most carefully engineered examples, it will be impossible to get free combination of possible optima from each.

Sophisticated construction is not necessary to bring out the across-candidate-set situation. The core of any voicing typology has nothing subtler than Faithfulness and Markedness playing off against each other.[4]

(1) **Primitive Voicing Typology 1.0**

Input	Outputs	M:voi	F:voi
1. da	a. da	*	
	b. ta		*
2. ad	a. ad	*	
	b. at		*

There are four 'grammatical systems' here — four combinations of output candidates, choosing one from each candidate set.[5] But only two systems are possible. The ranking conditions for each of the putative languages

System I	**System II**
⟨1a⟩ da→**d**a	⟨1b⟩ da→**t**a
⟨2b⟩ ad→a**t**	⟨2a⟩ ad→a**d**

are internally inconsistent. In System I , candidate ⟨1a⟩ needs F≫M and ⟨2b⟩ needs M≫F. In System II as well, we have opposite ranking requirements demanded by the candidates.

In the general case: perhaps *p* is available as an optimum in one candidate set and *q* in another: yet no grammar allows the two to coexist simultaneously, because they impose contradictory ranking requirements (which need not be as obvious as the ones just seen). Or optimal *p* may *demand* the appearance of *q*, ruling out otherwise viable alternatives, because the ranking requirements for the optimality of *p* entail those for *q*.

---

[2] Blowing things up beyond easy reckoning, serial theories admit *n*! formal grammars on *n* rules while typically recognizing vastly more rules than appear in any one grammar.

[3] Samek-Lodovici & Prince 1999, 2005 explore the first case; Samek-Lodovici 2001 deals with the second.

[4] Lombardi 1999 is the locus classicus and Ursprung of voicing typology work. We glance at her positional faithfulness system in next example.

[5] Four = free choice of one of 2 outputs from each of 2 candidate sets.

A slightly amplified voicing typology illustrates this latter situation. We add a constraint that demands faithfulness to voicing of those segments appearing in onset position in the output (Lombardi 1999, with roots in Beckman 1997).

(2) **Voicing Typology 2.0**

Input	Outputs	F/O:voi	M:voi	F:voi
1. da	a. da		*	
	b. ta	*		*
2. ad	a. ad		*	
	b. at			*

Missing now is only this system:

⟨1b⟩ da→**t**a
⟨2a⟩ ad→a**d**

Devoicing in *onset* implies devoicing everywhere, given these constraints. We have 6 formal grammars, but they do not even accommodate the 4 possible systems arrived at by choosing independently from possible optima.

A rich pattern of predictions about linguistic possibilities and impossibilities emerges almost inevitably from the way grammatical theory is set up. In general, we just can't have it all.

2. Why Patterns?

Here we look in more detail at the factors that frustrate vacuity.

2.1 Inside the Candidate Set

To get a sense of the within-candidate-set situation, we examine the simplest interesting case, a set of 3 constraints in which every ranking has its own distinct optimum:[6]

(3) **3 constraints**, 3! = 6 optima

/a/	**C_1**	**C_2**	**C_3**
a→ x_1	0	1	2
a→ x_2	0	2	1
a→ x_3	1	0	2
a→ x_4	1	2	0
a→ x_5	2	0	1
a→ x_6	2	1	0

With $C_1 \gg C_2 \gg C_3$ as the ranking, it is clear that ⟨a→x_1⟩ is the winner. Each permutation of ranking order supplies exactly one candidate with the winning violation profile (0,1,2). Since that candidate is different for each permutation, all six candidates are possible optima. Each wins under one ranking.

---

[6] This case is drawn from joint work with Vieri Samek-Lodovici.

This extremely symmetrical layout is not just an expository convenience; it's absolutely required to get the result. The key to the violation scheme is the number of discriminations between candidates — three — and not the exact numbers assigned. A constraint in this scheme could assign 2-4-6 violations, or 3-17-95, or indeed any triplet of distinct quantities.

Symmetry can be disrupted either among candidates or among the constraints. Some of the violation profiles might not designate linguistic structures. As an example of nonexistence, consider a mono-segmental input like /V/: there is no candidate with a violation score greater than 1 on the anti-deletion constraint MAX, even though the number of violations MAX can face is, over the set of all inputs, unlimited. In the case at hand, suppose that no candidate took on the violation profile (2,1,0). Then there are only 5 optima.[7] We can imagine a general violation space attached to a given set of constraints, in which each dimension corresponds to a constraint, and is big enough to accommodate the violation scores that its constraint assesses.[8] Ecologically, because of structural properties and interactions between them, any violation space is likely to be rather sparsely inhabited by forms.

A second candidate-based situation arises from presence rather than absence. The mere existence of certain candidates can interfere with others' potential for optimality. To take an extreme example, suppose there were a perfect candidate ⟨a→x_7⟩ which satisfied all the constraints. It will win under any ranking, and the violation profiles listed above, even if they describe linguistic forms, will be irrelevant. Example: the candidate ⟨CV→ $(CV)_\sigma$⟩ has this property in the basic CV syllable theory, guaranteeing that every language allows CV syllables, and that its competitors, in their multitudes, will languish forever unrealized.[9]

Within the *constraint* realm, shortfall will occur whenever a constraint, as a matter of definition, fails to make the requisite number of distinctions. Replace constraint C_3, for example, with a C_3' that assesses a maximum of *one* violation per candidate. Now things look like this:

(4) **3 constraints, 4 optima**

/a/	**C_1**	**C_2**	**C_3'**
a→x_1	0	1	1
a→x_2	0	2	1
a→x_3	1	0	1
a→x_4	1	2	0
a→x_5	2	0	1
a→x_6	2	1	0

To the unwary onlooker, everything looks much the same as before. No two candidates have the same violation profile, and there are still 6=3! of them. But the optimum structure has been crucially altered. Compare the first two candidates, constraint by constraint. The second candidate ⟨a→x_2⟩ is never better on any constraint than the first and it is worse on C_2.

Candidate ⟨a→x_2⟩ therefore cannot be optimal, because ranking has nothing to work with to make it so — no constraint that prefers it to ⟨a→x_1⟩. (This is a more interesting version of the situation induced by a perfect candidate.) In the lingo, it is said to be *harmonically bounded*, a

---

[7] In basic syllable structure, an input /CC/ has no output with 2 NOCODA violations, 1 DEPV violation, and 0 MAX violations. Getting 2 codas from this input requires 2 epentheses.

[8] If the full range of allowed constraint violations is seen in ex. (3), then its violation space is a cubical lattice where each dimension includes three points: 0,1,2.

[9] See OTCIGG:ch. 6, p.111, ex. (128) for discussion.

notion upon which we will expand in §7 below.[10] A similar fate has befallen the candidate ⟨a→x_5⟩, leaving us with only four potential optima out of six candidates.

A moment's thought will reveal that harmonic bounding cannot be some quirk that pops up occasionally to nix the odd form; rather, it is the common fate of nearly everything. Because generative linguistic theory admits structures of unbounded length and of unlimited structural elaboration, the number of candidates is sure to be unbounded as well. But the number of optima in any set is necessarily finite.[11] Consequently, *almost all* candidates are doomed by linguistic theory to be losers — harmonically bounded, never optimal, universally absent from human language.

We can think of the constraint set as a way of describing linguistic objects. To get the full range of optima from a given input, we need each constraint in a set of size *n* to be able to produce at least *n* descriptions. On the candidate-side, objects must exist that answer to the descriptions, and other better objects must fail to exist. Injecting reality into the matter inevitably compromises these conditions, resulting in systems with rich structure. We want to understand this structure.

2.2 *Optimum* vs. *optimum*

Cross-candidate-set incompatibilities arise when optima from different inputs impose contradictory ranking requirements. Here again, the effects are anything but sporadic or trivial. With true freedom of combination of optima from different input types, or anything remotely close to it, even the generosity of *n*! would be quickly overwhelmed. Commonly, implicational language acknowledges the limits: if a grammar admits *p*, then it must also admit *q*. This is always re-phrasable as incompatibility: the grammaticality of *p* rules out the alternatives to *q*.[12]

We've already seen simple examples drawn from voicing typology. Another of the same type, which has had an influence on the development of OT, comes from a hyper-basic syllable theory recognizing only ONSET, NOCODA, MAX, and DEP.[13] Let NOCODA and ONSET be undominated, so that only CV syllables appear in the output. Suppose that an input form /pat/, which must be unfaithfully realized, optimally suffers deletion rather than epenthesis:

pat→ $(\text{pa})_\sigma$
pat ↛ * $(\text{pa})_\sigma(\text{ti})_\sigma$

This means that we must have DEP≫MAX in order to privilege deletion of 'problematic' C, i.e. any input C whose presence entails a faithfulness violation.

A consequence now follows: problematic V must also necessarily be deleted. From /apo/, for example, we do not have the choice of epenthesis to supply an onset, to produce *e.g.* (ʔa)(po).

---

[10] The notion appears in Samek-Lodovici 1992 and in Prince & Smolensky 1993/2004:esp.209ff, where it is used extensively to show impossibility. It is explored in detail in Samek-Lodovici & Prince 1999, 2005. It is taken up in §7 below in relation to general entailment issues..

[11] Strictly speaking, it is the *violation profiles* that the theory of evaluation accounts optimal — it never sees the structures (Samek-Lodovici & Prince 2005). A violation profile may correspond to more than one structure.

[12] Because admitting P⇒Q 'P implies Q' is exactly the same as disallowing P&¬Q 'P and not-Q'.

[13] ONSET = 'No vowel-initial syllables'. NOCODA = 'No C-final syllables'. MAX = 'No deletion' (input is *max*imally realized). DEP = 'No insertion' (output *dep*ends on input). The last two originate as 'PARSE' and 'FILL' respectively. See OTCIGG:ch. 6,106ff., McCarthy & Prince 1995:12ff.

Instead, we get $(\mathrm{po})_\sigma$ whether we want it or not. Epenthesis here would require MAX≫DEP, exactly contrary to what's needed to obtain ⟨pat→pa⟩.

In this case, we have three possibly optimal outputs from /pat/ — pa, pat, pa.ti — and three from /apo/ — po, ʔa.po, a.po — yielding 3×3=9 combinations. But only 7 are possible linguistic systems in the hyper-basic theory. Such perfectly sound candidates as ⟨apo→ʔapo⟩ and ⟨pat→pa⟩ will never show up simultaneously in any grammar. Noting the lack of linguistic evidence for this particular linking of optima, Prince & Smolensky (2004:115) revise the constraint set so as to disentangle C-deletion/epenthesis from V-deletion/epenthesis.[14]

A number of more complex cases, some considerably so, are explored in OTCIGG under the heading of inventory structure, esp. in chs. 8 and 9. It is shown there, for example, that under a certain conception of place-of-articulation constraints, an effect of 'harmonic completeness' is obtained, whereby any inventory containing a complex articulation such as as labialized coronal must also contain simple labials and simple coronals (re-examined in §9.1 below). Other types of harmonic completeness emerge in a generalized syllable structure theory that responds to sonority distinctions (taken up in §9.2, §9.3 below). These have a Jakobsonian feel (structuralism being about the structure of inventories), but implicational phenomena are by no means restricted to collections of output forms. In the deletion/epenthesis example, the syllabic inventory is fixed at CV; what's at issue is how you can get there.

Output inventories *emerge* in systems which lack structuralist constraints delimiting their elements.[15] The action of a constraint hierarchy on the universal collection of candidate sets yields a grammatical system, pairing an output to each input. We may interest ourselves in the character of the outputs thus obtained, but the primary act of grammar is derivation, and any study of its properties must work through that.

3. Analyzing an Implication

A particularly striking case of implication and inconsistency within a realistic constraint system is examined in important recent work by Arto Anttila ("Variation in Optimality Theory," a collection of four handouts, henceforth VOT-1,2,3,4: 2006). He considers phenomena of consonant deletion and retention in American English, as analyzed in Kiparsky 1993. The effects show up in examples like these (VOT-1, ex. (7)):

(5)	It cost ~ cos' five dollars.	(t→∅ before C)
(6)	It cost ~ cos' us five dollars.	(t→∅ before V)
(7)	That's how much it cost ~ cos'##	(t→∅ before pause)

---

[14] The approach there is to divide the cognate of DEP into the equivalents of DEPC and DEPV (v. (57) below). The choice between V epenthesis and deletion (of C — this will be about disposing of input C) is determined by the relationship between MAX and DEPV; the choice between C epenthesis and deletion (of V, since this is about onsets) is determined by the independent relationship between MAX and DEPC. In the realm of basic syllable structure as delimited in OTCIGG, Ch. 6, the alternative is to distinguish MAXC and MAXV, or to make the same V/C distinction in both MAX and DEP, yielding interestingly different typologies.

[15] For discussion, see OTCIGG:154-156, 207ff

Of interest is not merely the fact of loss, but the relations that hold between loss in the various environments. Kiparsky cites the following (VOT-1: ex. (16)):

(8) **Implicational Universal**. If t,d-deletion occurs before a vowel or a pause, it also occurs before a consonant.

The idea is that variation arises because a speaker commands a number of different grammars. A major source of structure in variation is precisely the same kind of implications and inconsistencies that are found in any set of grammars.

Anttila notes that eyeballing a violation tableau does not rapidly deliver up the cited universal. Here is a representation of the assumed constraint set and its view of the candidates considered**:**[16]

(9) **Violation Table for the Constraint Set**

IN	OUT	*Complex	MWd-σ-init	Ons	Ph-MWd-fin	MaxC	*Remark*
1. cost us	a. cost.us	*		*			faith1
	b. cos.tus		*				faith2
	c. cos.us			*		*	***del***
2. cost me	a. cost.me	*					faith1
	b. cos.tme	*	*				faith2
	c. cos.me					*	***del***
3. cost##	a. cost.	*					faith
	b. cos.				*	*	***del***

The constraints, phrased informally, are these:[17]

(10) *Complex. No intrasyllabic C clusters.
(11) Ons. No vowel-initial σ.
(12) MWd-σ-init. The first element of a morphological (*i.e.* input) word begins an output syllable.[18]
(13) Phr-MWd-fin. The final element of an (output) phrase is the final element of an input word.[19]
(14) MaxC. No deletion of a consonant.

---

[16] To make sound arguments concerning optimality, the candidates considered must be sufficient to support the inference that *all others* are also bettered by the desired optimum. In this case, there must be many others that are competitive (for example, those with epenthesis), but we'll assume that they are eliminated by higher-ranked constraints (for example, DepV).

[17] Constraints (12) and (13) are framed after the Align of OTCIGG:127, later generalized into the Alignment family of McCarthy & Prince 1993a, and are named accordingly in the original. I have rearranged their names, and will shortly shorten them even further.

[18] This constraint militates against deletion of an underlyingly word-initial segment, syllabification across word boundary that pushes a underlying vowel out of word-initial position, initial epenthesis that does the same, etc.

[19] This constraint militates agains deletion of an underlying word final consonant when the word is in phrase-final position, insertion of a element phrase-finally, etc.

Perspicuity does not improve when we consider a ranked hierarchy and the optima it produces. Since we are interested in deletion patterns, let's look at a grammar that deletes everywhere. The ranking has been chosen and is given left-to-right; the optima are computed from that order.

(15) **Violation Tableau for a Ranked Hierarchy**

IN	OUT	*COMPLEX	MWd-σ-init	ONS	Phr-MWd-fin	MAXC	*Remark*
1. cost us	a. cost.us	* !		*			faith1
	b. cos.tus		* !				faith2
	c. cos.us			*		*	***del***
2. cost me	a. cost.me	* !					faith1
	b. cos.tme	* !	*				faith2
	c. cos.me					*	***del***
3. cost##	a. cost.	* !					faith
	b. cos.				*	*	***del***

Anttila (VOT-1: ex. 17) observes that the Implicational Universal (8) is not obvious from a violation tableau, nor even from the factorial typology, which he computes. Noting that "factorial typologies are hard for humans to understand" (VOT-1: ex. 21), he develops software which scans the typology to find implicational dependencies between optimal candidates in the various grammars (Anttila & Andrus 2006).

As valuable as such programs are for doing the heavy lifting, we explore another approach here. There is a reason for the imperspicuity visible in the displays given above: *they do not contain the relevant information*. Violation table (9) represents only the most basic individual performance of each candidate on each constraint and is generically suitable as a starting point for any theory of constraint violation. Tableau (15) takes place within OT proper and tells us what the optima are, given a certain ranking, and how the filtration of candidate sets proceeds. In neither case have we produced any information at all about what rankings are necessary or sufficient, given a prior choice of desired optima. Yet the inconsistencies and implications between optima are due entirely to the nature of the rankings that they each require.

Tableau (15) does contain an intriguing hint. All the candidates with word-final *t* retained syllable-finally are dismissed by *COMPLEX. But no calculation has been made that can tell us which constraints, if any, must be dominated by *COMPLEX to get this result, or whether there are other constraints that could stand in successfully for it.

We will find that the Implicational Universal (8) follows from the simple fact that the domination relation *COMPLEX≫MAXC is a necessary condition for deletion in each of the three cases, and it is *sufficient* for optimality in the pre-consonantal environment (case 2). Prevocalic and pre-pausal deletion each require it along with something more, so that pre-consonantal deletion follows as a concomitant of either (but not *vice versa*). A glance at the constraint set suggests the basic plausibility of this finding: how can we delete if MaxC is not dominated? what else besides *COMPLEX could compel cluster-reducing deletion? Our goal here is develop the tools to dig this kind of insight out of the data, to discern and deal with any loose ends, and to confirm its validity.

Brasoveanu and Prince (2005) usefully distinguish *finding optima given a ranking* (the 'selection problem') from *finding a ranking given the optima* (the 'ranking problem'). The logic of the two cases is quite different, as are the associated methods and procedures. To assault the 'the ranking problem' — delimiting the rankings required by some set of desired optima — it is only a first step to gather the individual violation profiles of the candidates (as in table (9)); and charting how the optima emerge under a given ranking (e.g. tableau (15)) is a distraction. We need explicit comparisons between desired optima and their competitors on each constraint.

The heart of ranking theory is the 'better than' relation. It is extended from individual constraints,where it simply means less violation, to an entire hierarchy, where ranking crucially comes into its definition as a means of adjudicating a possibly discordant collectivity of judgments. Candidate *q* is 'better than' candidate *x* on a *hierarchy* if and only if the highest-ranking constraint that distinguishes them favors *q* over *x*. Given this, the key to the ranking problem is to learn from *q* and *x* exactly which constraints can possibly stand in this crucial highest-ranked position, and which must absolutely be barred from it. This knowledge precisely delimits the set of rankings that are compatible with having *q* better than *x*. To acquire it, we must grasp how each constraint regards the *comparison* of *q* with *x*. Those constraints favoring *x* over *q* are particularly dangerous to the enterprise: they must be dominated, and dominated by at least one constraint that favors *q* over *x*. Constraints which see no difference between the two (because they incur the same number of violations) have nothing to say about the choice between them, and can go anywhere, as far as *q* and *x* are concerned.[20]

Raw violation-structure allows a constraint to discriminate an unlimited number of candidates, because violations start at 0 and may run on without limit. But *comparison* — OT — detects only three relations. Here, following Prince 2002a,b, we map them out, providing the (obvious) definition in terms of violation structure, writing C(*x*) for the number of violations C assigns to a candidate *x*, and '$\succ_C$' for the relation 'better than on constraint C'.

(16)	**Candidate relation**	**Violation Structure**	**Symbol**
	$q \succ_C z$	$C(q) < C(z)$	W
	$z \succ_C q$	$C(z) < C(q)$	L
	neither	$C(q) = C(z)$	*e*

The symbols become useful when we set out to calculate the comparative relations upon which ranking is based. They are based on the assumption that *q* is the desired ***w***inner, *z* the desired ***l***oser, in the contest between them.

With this is mind, let us introduce into the cells of the violation table (9) the relevant comparative information. Since we are seeking the rankings required to obtain some set of optima, they must be chosen in advance. Since we're interested in deletion patterns, let's focus on deletion in every case (yielding, overall, the delete-always grammar). For legibility, we omit the value '*e*'. The annotations indicate the outcome of each optimum *vs*. suboptimum comparison, and are placed in the suboptimum's row. Borrowing from Smolensky 2006, those marks incurred by the desired optimum are circled for clarity.

---

[20] This summary paraphrases OTCIGG:129-130.

(17) **Comparatively Annotated Tableau: Deletion everywhere**

IN	OUT	*CPLX	MWd	ONS	PHR	MAXC	*Remark*
1. cost us	a. cost.us	* **W**		*		**L**	faith1
	b. cos.tus		* **W**	**L**		**L**	faith2
	c. cos.us			⊛		⊛	***del***
2. cost me	a. cost.me	* **W**				**L**	faith1
	b. cos.tme	* **W**	* **W**			**L**	faith2
	c. cos.me					⊛	***del***
3. cost##	a. cost.	* **W**			**L**	**L**	faith
	b. cos.				⊛	⊛	***del***

The annotations are developed as follows: each suboptimum cell is compared against the corresponding cell of the desired optimum. If the desired loser's cell contains ***l**ess* violation, so that it actually does better in the confrontation than the desired winner, an L is planted in the its cell. If the desired winner betters the desired loser, by virtue of ***w**orse* violation in the loser's cell, a W is inscribed there. The desired optimum's violation profile remains unannotated; it is merely a kind of yardstick against which each competitor is measured, constraint by constraint. The generic results of this procedure are spelled out in the following summary, where the nonnegative variables a,b,c represent numerical quantity of violation; we write h_i for strictly positive constants.[21]

(18) Annotating a violation tableau

desired optimum	a	$b+h_2$	c
competitor	$a+h_1$ **W**	b **L**	c (***e***)

Analysis reveals more if we omit more. The pair-wise comparisons carry the ranking information. We tabulate exactly and only these, erasing the marks, whose role has been exhausted. We name the comparative rows after the suboptimum that generates them, so that [1a] denotes the result of comparing ⟨1c⟩ with ⟨1a⟩.[22]

(19) **Comparative Tableau: Deletion everywhere**

IN	OUT:**W**inner *vs.* **L**oser	*CPLX	MWd	ONS	PHR	MAXC	*Remark*
1. cost us	a. **cos.**us ≻ cost.us	**W**				**L**	***del*** ≻ faith1
	b. **cos.**us ≻ cos.tus		**W**	**L**		**L**	***del*** ≻ faith2
2. cost me	a. **cos.**me ≻ cost.me	**W**				**L**	***del*** ≻ faith1
	b. **cos.**me ≻ cos.tme	**W**	**W**			**L**	***del*** ≻ faith2
3. cost##	a. **cos.** ≻ cost.	**W**			**L**	**L**	***del*** ≻ faith

---

[21] Evaluation never sees these integers: only the greater than/less than relation, i.e. the sign of the difference in violation score between two competitors. See OTCIGG:ch. 5 for a method of reckoning directly with the multisets of violation marks which never comes near their cardinalities. (A 'multiset' is a collection of items with repetitions. In OTCIGG, each constraint produces a multiset of 'marks' when faced by a candidate. Formally, a multiset is a set equipped with a function that gives the muliplicity of each element.)

[22] This concise method of reference works fine as long as it's clear which candidate is the desired optimum, as in the present case. When this becomes an issue, we expand to, e.g., [1ca], which denotes the result of comparing desired optimum ⟨1c⟩ with competitor ⟨1a⟩.

What do the individual rows tell us? Let's focus on those rows that relate *Cplx and MaxC, all but [1b]. (We abbreviate the cases as follows: 'pre-V' = prevocalic; 'pre-C' = pre-consonantal; 'pre-P' = pre-pause, phrase final.)

(20) **ERCs involving *Cplx**

ERC#	*Cplx	MWd	Ons	Phr	MaxC	*Remark*
[1a]	**W**				**L**	pre-V: **del** ≻ faith1
[2a]	**W**				**L**	pre-C: **del** ≻ faith1
[2b]	**W**	**W**			**L**	pre-C: **del** ≻ faith2
[3a]	**W**			**L**	**L**	pre-P **del** ≻ faith

An immediate observation: the first two rows are *identical*. This shared condition yokes the optima from the 1st and 2nd candidate sets together. Pre-V deletion is linked here, quite obviously, to pre-C deletion.

The full tableau (19) indicates that their correlation is incomplete, because each faces a separate additional requirement to achieve optimality — row [1a] has [1b] and row [2a] has [2b] as accompaniment. But the pre-C situation is simpler than it seems. Satisfaction of both [2a] and [2b] is indeed necessary, but the two rows do not express independent requirements. Row [2a] contains all the information, as we will see shortly. From this it follows that the presence of pre-V deletion, which delivers [1a], the clone of [2a], will guarantee that pre-C deletion takes place as well.

This is already half of the Implicational Universal (8). To fill in the details of the argument, consider the *meaning* of the pre-C-deletion comparisons [2a] and [2b].

Row [2a] has the form (W,e,e,e,L). If the desired winner is to win, any decisive highest-ranked constraint must proclaim via W its winner-oriented preference. The L-constraint must be dominated, or else it could occupy the place of the highest-ranked-distinguisher and thereby upset the preferential applecart. Since only one W-constraint is available to do the dominating, we may safely conclude that it *must* dominate. Ranking-wise, we have:

(21) **Content of [2a]**: (W,e,e,eL)
***Cplx ≫ MaxC**

By constrast, ERC [2b] is of the form (W,W,e,e,L) and supplies *two* potential dominators for MaxC. Either one will do, as far as this row goes, and we can only conclude this from it:

(22) **Content of [2b]**: (W,W,e,e,L)
Cplx ≫ MaxC *or MWd≫MaxC

It is apparent from (22) that [2b] is merely a weakened version of [2a]. If, as in any pre-C deleting grammar, we have the [2a] ranking, we will also have trivially satisfied [2b] . Row [2a] says it all, and since it is literally identical with [1a], it follows as promised that pre-V deletion entails pre-C deletion.

Half-done is well-begun: let's turn now to the other part of the Implicational Universal, which relates the pre-pause to the pre-C case. Here again there is a direct pay-off for examining the ranking content of the relevant comparative rows. In the data at hand, pre-pause deletion reduces to the satisfaction of the single ERC [3a], which has the form (W,e,e,L,L).

● The 2nd- and 3rd-listed constraints in the row (W,e,e,L,L), MWd and ONS, shoot blanks and do not participate in the choice between *cos'* and *cost* in phrase-final position.

● The last two constraints in the list, PHR and MaxC, favor the desired loser (L) and therefore present an immediate danger to the desired winner. Both must be dominated.

● Only the 1st-listed *CPLX favors the desired winner (W) , so it alone provides a viable dominator. Conclusion: *CPLX must dominate *both* MAXC and PHR.

(23) **Content of [3a]**: (W,e,e,L,L)
CPLX≫MAXC *and *CPLX≫PHR

Row [3a] requires that two separate ranking relations hold. One of the conjuncts is the same as what's required by [1a] and its twin [2a]. Therefore, in any pre-pausally deleting grammar, it must also be the case that deletion occurs before C. The Implicational Universal has been parsed.

(24) **Implicational Universal** (repeated). If t,d-deletion occurs before a vowel or a pause, it also occurs before a consonant.

The core of the matter, now clear, is the relation *CPLX≫MaxC. This is hardly surprising, in broad terms, since the phenomenon turns on the simplification of complex clusters by deletion. But it is only from scrutiny of the details that the implicational argument emerges.[23]

In sum, the argument runs like this. Any deletion of a consonant from a complex cluster needs *CPLX≫MAXC. (This is established in the data by [1a],[2a],[3a].) Prepausal deletion requires it, as does pre-V deletion; but each requires something else in addition. By contrast, pre-C deletion requires no more than this — it is sufficient as well as necessary. Therefore, the occurrence of deletion either pre-V or prepausally will necessarily entail pre-C deletion, though not *vice versa*. The Universal Implication follows from a single markedness-faithfulness relation. All of this emerges quite straighforwardly from the comparative tableau.

There is more. Anttila finds other significant implicational relations lurking in the constraint system (VOT-1:ex 23), unremarked in the original discussion. He notes, in particular, these two concerning *retention* of the final cluster in word-aligned position, as in cost→ cost)$_{\sigma}$. They are re-phrased here.

(25) **Aligned Retention-1**. If t,d are retained as word-aligned before V, then they are so retained before C.

*cos***t***.us* ⇒ *cos***t***.me*

(26) **Aligned Retention-2**. If t,d are retained as word-aligned before C, then they are so retained before pause.

*cos***t.***me* ⇒ *cos***t.**##

Unfolding the causes requires the same kind of analysis we've undertaken so far. It is useful, then, to pull together what we've seen and set our tools in order.

---

[23] Note that 'scrutiny' here does not need a high-powered logical instrument to be effective: it's more like turning the head to look in the right direction.

4. The Elements

In the simplest case, a comparison based on a desired ‘better than’ relation between two candidates *q*, *x* — aiming to set up a constraint hierarchy that yields *q*≻*x* — produces an outcome like (W,e,e,e,L), as in (21) above. Here there is no choice but to require that the first-listed constraint dominate the last. More complicated cases involve multiple W’s or L’s. A general principle, derived directly from the definition of ‘better than’, interprets any comparative tableau row: *some* W must dominate *all* L’s.

We have seen this at work concretely in the examples just explored. Recall the pre-C competion [2b] between deletion and *mis*-aligned retention, as in *cos***’***.me* and *cos.***t***me*.

(27) **C deletes before C**

IN	OUT:**W**inner *vs.* **L**oser	*CPLX	MWd	ONS	PHR	MAXC	*Remark*
2. cost me	b. **cos.**me ≻ cos.**t**me	**W**	**W**			**L**	pre-C: ***del***≻faith2

The relevant comparative profile runs (W,W,e,e,L). One of the W’s must dominate the L in any grammar that succeeds in making *cos.me* better than *cost.me*. If not, then the L-constraint will stand in the ‘decider’ position, reversing the desired ruling.

Contrast now the comparison of prepausal deletion with aligned retention, [3a]:

(28) **Comparative Tableau: Deletion before Pause**

IN	OUT:**W**inner *vs.* **L**oser	*CPLX	MWd	ONS	PHR	MAXC	*Remark*
3. cost##	a. **cos.** ≻ cos**t**.	**W**			**L**	**L**	pre-P: ***del*** ≻ faith

The comparative profile is (W,e,e,L,L). The W must dominate both L’s — if either is allowed to dominate, the retentive candidate gains an unwelcome local victory.

The interpretive rubric “some W must dominate all L’s” is sufficiently important to merit a name: the Elementary Ranking Condition (ERC).

(29) **Elementary Ranking Condition (ERC)**. Given a comparison of two candidates *q*, *x* over a constraint set Σ, a hierarchy (linear ordering) H on Σ will obtain the result *q*≻*x* iff H satisfies the following ‘elementary ranking condition’: every constraint assessing[24] L of (*q*,*x*) is dominated by some constraint assessing W of (*q*,*x*).

We will write [*q*≻*x*] for the ERC associated with the pair (*q*,*x*). To chop the basic logic:

- **Validity**. If there are no L’s in [*q*≻*x*], it is vacuously true, logically valid, so that any ranking satisifies it. No constraint need be dominated to get the result. One interesting subcase occurs when *q* and *x* have the same *violation* profile; the ERC is all *e*’s, and [*q*≻*x*] is the same as [*x*≻*q*]. The ERC [*q*≻*x*], then, really determines the conditions under which *q* is better-than *or* just-as-good-as *x*. It might be better to write [*q*≿*x*], but we shirk the tittle.[25]

---

[24] To spell out ‘assessing a comparative value of an ordered pair’: a constraint assesses W of any such pair, if it evaluates the first as better than the second; assesses L if the second is evaluated as better than the first; and *e* if the result is ‘nothing, neither way’.

[25] The standard notation is [*q*~*z*]. Here we choose to emphasize the quest for betterness.

● **Invalidity**. When there are some L's, but no W's, the ERC is logically invalid, and its requirements can never be met. Every L constraint must be dominated, but there's no constraint fit to do the job.

ERCs which are logically valid or logically invalid will be called 'trivial', the others (with at least one L and at least one W), nontrivial. Of these, we notice the following:

● ∀∃ *vs.* ∃∀. When there are both L's and W's in an ERC, the statement 'every L is dominated by some W' becomes equivalent to the rubric we began with: 'some W dominates every L'. Since a formal hierarchy is a strict linear order, there must be one W-constraint that is highest-ranked among those under consideration.

The set of ERCs associated with a collection of desired optima contains *all* the ranking information inherent in the candidate sets under consideration. As we have seen, an ERC set may display considerable internal structure. Of particular practical importance are the relations of entailment and inconsistency within an ERC set. Fortunately, these are easily determined from the representation of the ERC as a list (or 'vector') of comparative values.[26]

The theory of entailments from a single ERC proves to be remarkably simple and can be completely understood in terms of relations among corresponding W,L,*e*-values (Prince 2002a: 5-7). Beyond the blindingly obvious observation that entailment is not threatened when corresponding values are identical, the relations are just two in number.

(30) **W-extension**. Replace any value in an ERC with a W, and the result is entailed by the original.

We have seen an example: (W,W,e,e,L) follows from (W,e,e,e,L) by W-extension in 2nd position.

W-extension works because it adds a disjunct to the ranking condition, weakening it. The ERC (W,e,e,e,L) says $C_1 \gg C_5$. the W-extended version (W,W,e,e,L) offers *either* C_1 or C_2 as the crucial dominator.[27]

The second relation involves L:

(31) **L-retraction**. Remove an L from any ERC, and the result is entailed by the original.

We have already worked with an example: (W,e,e,L,L) entails (W,e,e,e,L) by L-retraction in the penultimate position.

L-retraction works because it removes a conjunct from the ranking condition, weakening it.[28] The ERC (W,e,e,L,L) asserts that both of the last two constraints must be dominated; the L-retracted version (W,e,e,e,L) demands only the domination of last constraint and has nothing to say about the second to last.

Entailment thus runs along the course L→e→W, with each move along the path resulting in a weakening of ranking demands.

---

[26] Strictly speaking, we ought to maintain a distinction between the ERC — a logical expression — and the list of W,L,e values, the ERC *vector*. Since they stand in an unambiguous 1:1 relation, we'll use them interchangeably, risking little in the way of meaningful confusion.

[27] As the propositional calculus puts it, $P \Rightarrow P \vee Q$.

[28] Talking prop calc: $P \& Q \Rightarrow P$.

The C-deletion tableau is worth another look, with this perspective in mind. We focus, as above, on those constraints establishing relations between between *Cplx and MaxC.

(32) **Entailment structure**: [3a] ⇒ [1a,2a] ⇒ [2b]

ERC#	*Cplx	MWd	Ons	Phr	MaxC	Remark
[3a]	**W**			***L***	**L**	strongest
[1a] = [2a]	**W**				**L**	L-retraction
[2b]	**W**	***W***			**L**	W-extension

It is immediately clear that the first row entails all the others. The second and third row are L-retracted from it. The last row is W-extended from the second, and is therefore entailed by both preceding rows.

It's worth noting that each of these observations has a slightly different effect on the flow of the argument. The within-candidate-set relation [2a]⇒[2b] allows us to *ignore* [2b] and rest the optimality of pre-C deletion entirely on satisfaction of [2a]. The cross-candidate-set relations [1a]⇔[2a] and [3a]⇒[1a,2a] yield the Implicational Universal (24).[29]

Anttila's further implications are concerned with *retention* rather than deletion. By Aligned Retention-1, the optimality of *cost.us* with the pre-V environment implies that /cost/ comes out faithfully and word-aligned in all contexts. By Aligned Retention-2, regardless of the fate of the pre-V form, optimal *cost.me* entails optimal phrase-final *cost*. Examining the Aligned-Retention grammar is a natural move (see Appendix, p.59.), but unnecessary in this case. Appearances notwithstanding, these implications too lie amid the ERCs arrayed before us.

The key is to derive further ERCs from the ones we have. If we reverse the order of comparison, transforming the evaluation of $q \succ x$ into evaluation of $x \succ q$, for violationwise-distinct candidates *q* and *z*, we produce the logical *negation* of the original ERC (Prince 2002a:12ff.).[30] Swapping winner and loser in the comparison has the effect of switching W and L in the associated ERC, leaving *e* untouched. The ERC produced by winner/loser reversal is called the 'negative' of the original, and we can write ¬α for this entity.

The identity of the desired optimum has been notationally suppressed in the ERC naming convention used until now: we have written, for example, [1a], depending on context to remind us that candidate ⟨1c⟩ is the output desired optimal, rather than ⟨1b⟩. Since the reversal operation crucial manipulates this suppressed information, we expand the naming convention: instead of [1a], we will write [1ca], which refers to the input ('1'), the desired optimum ('c'), and its competitor ('a').

---

[29] If we are only interested in the minimal set ranking relations required to get the system, a 'basis' in the terms of Brasoveanu and Prince 2005, we can also dispense with [1a] and [2b]. Here we are looking at implications between optima coming from different candidate sets , so it is valuable to retain them. The fact that candidate set 2 contributes nothing to the basis for the system necessarily means that its conditions are entailed.

[30] NB: $q \neq x$. Reversal negates in all cases except the 'degenerate' one in which x and y are violation-wise identical, so that the ERC consists entirely of *e*'s (Prince 2002a:13). The logic of ERCs is nonclassical (it is the implication-negation fragment of RM3, Prince 2002a:47ff, Anderson & Belnap 1975, Meyer 1975, Parks 1972, Sobociński 1952), and admits a case where ¬p and p both hold, namely when each has the value *e*. In ERC terms, we can have both $[q \succ x]$ and $[x \succ q]$ holding, just in case *q* and *x* have identical violation profiles. Because of this effect, '$[q \succsim x]$', which we have resisted, again recommends itself as a notation.

ERC [1ca] is concerned with the relation '***cos.***_us_ ≻ *cost.us*'. It compares deletion against aligned retention, pre-V. Its negative [1ac] evaluates '***cost.***_us_ ≻ *cos.us*'. This gives a necessary condition for *retaining* final *t* (with the given syllabification) in the pre-V environment. It is clearly inconsistent with ERC [2ca] = [1ca] (evaluating '*cos.me* ≻ *cost.me*'), which supplies a necessary and indeed sufficient condition for pre-C deletion.

(33) **Reversing [1ca]**

ERC#	*CPLX	MWd	ONS	PHR	MAXC	Remark
¬[1ca]	**L**				**W**	pre-V: faith ≻ del
[2ca]	**W**				**L**	pre-C: del ≻ faith

The inconsistency arises because the first row demands MAXC≫*CPLX and the second row demands the exact opposite.

Reversing [1ca] clearly requires reversing the identical [2ca] to get a consistent grammar. Therefore, if 'cost.us' is optimal, it must also be the case that at least one of the conditions for optimal 'cost.me' is met, namely ¬[2ca]. What are the others? It is worthwhile to recall the entirety of candidate set #2.

(34) **Pre-C Candidates**

IN	OUT	*CPLX	MWd	ONS	Phr	MAXC	*Remark*
2. cost me	a. cost.me	*					faith1
	b. cos.tme	*	*				faith2
	c. cos.me					*	del

The ERC we are calling ¬[2ca] relates candidate ⟨a⟩ as desired optimum to ⟨c⟩. But the optimum must also defeat ⟨b⟩. A quick check shows that ⟨b⟩ is harmonically bounded by ⟨a⟩:

(35) **Harmonic bounding check**

IN	OUT	*CPLX	MWd	ONS	Phr	MAXC	*Remark*
2. cost me	a. cost.me	⊛					faith1
	b. cos.tme	*	* **W**				faith2

Candidate ⟨b⟩ is eliminated under any ranking, and may be dropped from consideration. Only one ERC need be satisfied to ensure the optimality of *cost.me*, and that is ¬[2ca]=[2ac], which compares aligned retention with deletion. As we've seen, we obtain this very ERC from pre-V retention, as ¬[1ca]. Aligned-Retention-1 (25) follows, the first new implication, repeated here.

(36) **Aligned-Retention-1**. If t,d are retained as word-aligned before V, then they are so retained before C. Output-wise, *cos***t***.us* ⇒ *cos***t***.me*.

Now consider the effect of retaining the *t* before C. This requires '*cos***t***.me* ≻ *cos.me*', which is precisely the ERC ¬[2ca]. As we can see below, it is *not* consistent with phrase-final deletion:

(37) **Reversing [2a]**

ERC#	*CPLX	MWd	ONS	PHR	MAXC	Remark
¬ [2ca]	**L**				**W**	pre-C: faith ≻ del
[3ba]	**W**			**L**	**L**	pre-P: del ≻ faith

ERC ¬[2ca] requires MAXC≫*CPLX. But the opposite relation *CPLX≫MaxC is one of the conjuncts of [3ba]. The inconsistency tells us that pre-C retention must lead to phrase-final retention. This may also be directly put as entailment:

(38) **Reversing [3a]**

ERC#	*CPLX	MWd	ONS	PHR	MAXC	Remark
¬[2ca]	**L**				**W**	entails next row
¬[3ba]	**L**			**W**	**W**	W-extended

By W-extension, it is clear that ¬[2ca] entails ¬[3ba].[31] Since each is sufficient for the achievement of its own desired optimum, we have derived Aligned Retention-2, repeated here from (26):

(39) **Aligned Retention-2**. If *t*,*d* are retained as word-aligned before C, then they are so retained before pause. Output-wise, *cost.me* ⇒ *cost.##*

Once again the implication is driven by the relation between MAXC and *CPLX, now in the form of MAXC≫*CPLX, and by nothing else. This ranking is necessary for Aligned Retention before V and C (though not before pause), and sufficient for Aligned Retention before C and pause. Hence the result.

It is no offense to intuition to discover this — after all, these two are the basic markedness/faithfulness constraints involved in the phenomenon at hand, which is all about simplifying or retaining a complex cluster. It is perhaps unexpected that the implicational structure follows so directly from their relation. Beyond that, and authentically surprising to at least one spectator, is the fact that the further implications discovered by Anttila are, in the Kiparsky system, *logically equivalent* to the original Implicational Universal – and thus, in an abstract sense, not new at all.

The ERCs involved are [1ca],[2ca] and [3ba]. The logical relations [1ca]⇔[2ca] and [3ba]⇒[2ca] deliver the first Implicational Universal. The further implications follow because ¬[1ca]⇔¬ [2ca] and ¬[2ca]⇒¬[3ba]. From the logical point of view, we have a mere restatement, not the introduction of new information. This provides a rather dramatic illustration of the way that theory uncovers structure in data — and the way that data-structure becomes visible when we look for structure in theory.

In the investigation, the crucial facilitating tool was the ERC, and ERC entailment was the crucial logical relation that revealed the relations in the data.

ERC entailment, we found, depends on the relations that are incorporated into the notions of W-extension and L-retraction. From the suboptimal competitor's point of view, the values 'W','e', and 'L' reflect the ability of a constraint, when ranked, to dismiss it from consideration when the constraints are ranked. 'W' marks the existence of a comparative flaw in the desired loser that can decide in favor of the desired winner; 'L' marks an advantage in the loser that could tilt the decision in its favor. We have entailment, putting it broadly, when the entailed ERC's loser shows either *more* such flaws (W-extension) or *fewer* such advantages (L-retraction) with respect to its desired winner. (We use the loose terms 'more' and 'fewer' with set-theoretic inclusion in mind, intending superset and subset relations, respectively.)

---

[31] This is entirely as we'd expect on logical grounds, given that ¬[2a]⇒¬[3a] contraposes [3a]⇒[2a].

Turning things around, ‘W’ marks a comparative advantage enjoyed by the desired optimum, which can lead to its victory in the competition, and ‘L’ marks a comparative flaw in it that can lead to its demise. It should be clear that if one desired winner has a subset of the occasions for victory that another one has, or a superset of the occasions for defeat, then the success of the first (under its less favorable conditions) will ensure that the second one also succeeds.

In sum: If success can be achieved against the problems facing ERC α, then when β has fewer problems or more means for overcoming them, as guaranteed by the truth of $\alpha \Rightarrow \beta$, then surely β will also succeed.

5. Against interpretation

Resolving the Anttila/Kiparsky implications led us to look for inconsistency in sets of comparative data. Our technique was to *interpret* various ERCs in terms of ranking and then argue — like schoolmen parsing Aristotle in Latin translation — that the ranking conditions thus derived could not be simultaneously sustained. Such translation and logic-chopping turns out to be unnecessary. ERCs themselves support an operation that delivers the result without the interpretive detour.

From two ERCs we can construct a third, their *fusion*, by combining corresponding values according to a scheme which closely resembles the truth table for logical conjunction. Take ‘W’ to be like ‘T’ and ‘L’ to be like ‘F’. Their combination runs accordingly:

(40) **Combining Values**

Fusion	Conjunction
$W \circ W = W$	T&T = T
$W \circ L = L$	T&F = F
$L \circ W = L$	F&T = F
$L \circ L = L$	F&F = F

ERCs go beyond the binary language of Boole in recognizing a third value: *e*. To achieve a minimal extension of conjunction that acknowledges the opinion of one conjunct when the other has none, we treat *e* as an identity element.

(41) Identity

$W \circ e = e \circ W = W$

$e \circ L = L \circ e = L$

$e \circ e = e$

Fusion retains much of the good behavior of conjunction as an operation: it is associative and commutative, so the order of fusing items in larger collections makes no difference to the outcome, and we may speak of the fusion of a *set* of ERCs, not just a pair. The fusion is a single ERC, which often contains valuable (if sometimes incomplete) information about the content of the set that was fused.

To see the effect of fusion, recall the simplest case of inconsistency from (33):

(42) Reversing [1a]

ERC#	*CPLX	MWd	ONS	PHR	MAXC	Remark
¬[1ca]	**L**				**W**	pre-V: faith ≻ del
[2ca]	**W**				**L**	pre-C: del ≻ faith
¬[1ca]∘[2ca]	**L**				**L**	*fusion of the two*

The inconsistency results in a fused row that contains no Ws. It is evident that this row cannot be satisfied. By the definition of ERC, all L’s must be dominated by some W. But there is no W to shoulder the task.

This is an entirely general phenomenon. If a set of ERCs is inconsistent — if there is no ranking that will satisfy all the conditions it imposes — then it contains a subset that fuses to a single row containing noWs. In the jargon, such rows are said to ‘fuse to L^+, where L^+ is the set of all W-free ERC rows containing at least one L. This is analogous to the appearance of all F’s in the truth table of a boolean logical contradiction like P&¬P, and to the fact that an inconsistent set of prop calc wffs like {P, ¬P}conjoins to a single formula that is F under every valuation. The third value introduces the twist that we are only guaranteed that a *subset*, possibly proper, fuses to L^+. To see this, recall that in our example, we also must have an ERC derived from loser ⟨1b⟩, output ‘cos.tus’, now running against retentional ⟨1a⟩.

(43) **Remembering ⟨1b⟩**

ERC#	*CPLX	MWd	ONS	PHR	MAXC	*Remark*
[1ab]	**L**	***W***	***L***			pre-V: *misaligned retention
¬[1ca]	**L**				**W**	pre-V: *deletion
[2ca]	**W**				**L**	pre-C: del ≻ retention
fu-all	**L**	***W***	***L***		**L**	*fusion of all three*

Now fusion of the whole *does not* lie in L^+. But the set is just as inconsistent as it was before, because no ranking exists that simultaneously satisfies everything in it. Adding something to an inconsistent collection of statements cannot make it consistent.

There is an easy procedure for testing the consistency of any ERC set: simply discard any ERCs that deposit W in the fusion of the set. They cannot possibly be members of a subset fusing to L^+. Repeat the procedure on the remaining ERCs, if any. And repeat again until there are no ERCs left (consistency), or until a subset fusing to L^+ is reached (inconsistency).[32] In the case at hand, if we start out with (43), we observe that [1ab] is guilty of placing the W in the fusion; removing it, we have left only ¬[1ca] and [2ca], which fuse to L^+.

The meaning of inconsistency is that the constraint set cannot supply a grammmar for the data: no ranking works. In the course of linguistic analysis, when we know that the data is right, this can indicate that something is wrong with the constraint set — perhaps a constraint needs to be redefined, perhaps a necessary constraint has not been included or discovered. But it may also diagnose a problem or a pattern in the data. As we’ve seen, the inconsistency of ¬[1ca] and [2ca]

---

[32] This procedure, drawn from Prince 2002b, mirrors and simplifies Recursive Constraint Demotion, RCD (Tesar & Smolensky 1993, 2004, Prince 2002a,b; Brasoveanu & Prince 2005). In RCD, the stages of the process are remembered as ordered strata of constraints, which yield a ranking that satisfies the ERCs iff satisfaction is possible.

means that pre-V retention (*cost.us*) is incompatible with pre-C deletion (*cos.me*). This is not a failure of descriptive success, but a prediction of considerable interest. And whether the outcome is desirable or disgruntling, the translation of observed or desired candidate relations into constraint relations via the ERC is what drives the investigation.

Because of the intimate relation between inconsistency and entailment, the fusion operation allows us to complete the theory of ERC logic. So far we have only considered the case of entailments arising from a single ERC: W-extension and L-retraction exhaust its consequences. With fusion in hand, we can comprehend the general case, in which an arbitrary *set* of ERCs is the basis for entailment. The generalization is that any ERC that follows from an ERC set also follows from a *single* ERC that is the fusion of one of its subsets. W-extension and L-retraction, taken with fusion, are the only tools required to extract the nontrivial consequences from a consistent set of ERCs.[33]

Crucial entailments arise in quite ordinary contexts. Consider these ERCs:

(44) **An implication from transitivity**

	C_1	C_2	C_3
α	**W**	**L**	e
β	e	**W**	**L**
α∘β	**W**	**L**	**L**

Direct relations are established only between the C_1 and C_2 (ERC α) and between C_2 and C_3 (ERC β). The natural question is, then, whether anything follows about the relation between the C_1 and C_3, which are not related by any single ERC. Fusion gives the answer:

$$\begin{aligned} \alpha\circ\beta = {} & (W,L,L) \\ & (W,L,L) \Rightarrow (W,e,L) \\ & \qquad\qquad\text{“}C_1 \gg C_3\text{”} \end{aligned}$$

Conclusion: C_1 must dominate C_3 in any grammar satisfying α and β, because α∘β tells us so.

Lest it be imagined that all such entailments are as obvious as this one, and therefore scarcely worthy of note, consider the following quite similar-looking case:

---

[33] The caveats *nontrivial* and *consistent* show up because W-extension and L-retraction don't give all the trivial consequences of an ERC, nor do they give all the consequences of a trivial ERC. For example, (e,W) doesn't W-extend to (W,e), but each booleanly entails the other. Similarly, (L,e) doesn't L-retract to (W,L), although as an invalid ERC, it surely entails it by the boolean apothegm *ex falso quodlibet*. The logic of ERCs (RM3, a relevance logic) makes distinctions within the valid and the invalid that boolean logic knows not of (Prince 2002a:47ff). In the case at hand, for example, (e,W) tells us not only that the desired candidate relation is satisfied in every ranking, it also identifies the very constraint that decides the matter. From the discriminating point of view of RM3, (e,W) therefore differs in entailments from (W,e), though they are both indistinguishably valid in the two-valued world.

(45) **Disjunctions galore**

	C_1	C_2	C_3
α′	**W**	**L**	**W**
β′	**W**	**W**	**L**

Ask again: Does this mean that $C_1 \gg C_3$?

ERC β′ tells only that C_1 *or* C_2 dominates C_3. If we interpret the entire tableau into boolean logic, via the definition of ERC, we get this tangle:

$$(C_1 \gg C_2 \textbf{\textit{ or }} C_3 \gg C_2) \textbf{ and } (C_1 \gg C_3 \textbf{ or } C_2 \gg C_3)$$

I leave it to reader, and the distributive laws taken together with the transitivity and asymmetry of '≫', to untie this knot. Fusion gives the same result as above,

$$\alpha' \circ \beta' = \alpha \circ \beta = (\mathrm{W,L,L})$$

so that, as in (44), we are immediately licensed to conclude that the answer is affirmative.[34]

The tools assembled here also allow us to resolve the problem of *independence*. In the consonant-deletion case, we'd like to know for sure that ERC [1cb] is *not* entailed by the others. In short — do we really need it?

(46) **All the ERCs for the All Deletion Grammar**

ERC#	*Cplx	MWd	Ons	Phr	MaxC	*Remark*
[1ca]=[2ca]	**W**				**L**	pre-V: **del** ≻ faith1
[2cb]	**W**	**W**			**L**	pre-C: **del** ≻ faith2
[3ba]	**W**			**L**	**L**	pre-P **del** ≻ faith
[1cb]		**W**	**L**		**L**	pre-V **del** ≻ faith2

The key observation is that P⇒Q holds precisely when P&¬Q doesn't. To test for the entailment P⇒Q is just the same as testing for the *inconsistency* of P&¬Q. We want to know whether the other ERCs entail [1cb]. This is the same as asking whether they are inconsistent with ¬[1cb].[35]

(47) The ERC set with ¬[1cb] substituted in for [1cb]

ERC#	*Cplx	MWd	Ons	Phr	MaxC	*Remark*
[1ca]=[2ca]	**W**				**L**	pre-V: **del** ≻ faith1
[2cb]	**W**	**W**			**L**	pre-C: **del** ≻ faith2
[3ba]	**W**			**L**	**L**	pre-P **del** ≻ faith
¬[1cb]		**L**	**W**		**W**	pre-V **del** ≻ faith2
fu-all	**W**	**L**	**W**	**L**	**L**	*fusion of all ERCs*

---

[34] Note that $\alpha' \circ \beta' = \alpha \circ \beta$, even though the two systems are not equivalent. The {α,β} system requires $C_1 \gg C_2 \gg C_3$ but the {α′ ,β′ } system only forces $C_1 \gg \{C_2, C_3\}$. Fusion can extract useful information even while losing some. This contrasts with boolean conjunction, which retains all information. We can't use conjunction and stay within ERC territory because it is often the case that the logical conjunction of two ERCs does not itself correspond to an ERC. (For example, the logical conjunction of α and β says "$C_1 \gg C_2$ *and* $C_2 \gg C_3$," which is not syntactically viable as an ERC, because C_2 is described as both a dominator and a dominee.) Fusion is as good as it gets in ERC-ville.

[35] See Prince 2002a:13, ex. (23).

Let's run the inconsistency test.

- fuse(all) = (W,L,W,L,L)

The top three ERCs each deposit a W in the first position (*CPLX). The bottom ERC deposits a W in the third position (ONS). Removing these ERCs leaves nothing behind. There is no subset that fuses to L^+.

- Conclusion: this set of ERCs is consistent.

ERC [1cb] is, then, *not* entailed by the others. It deals with a situation that is independent of the issues raised in the other data, and can thus join in happily with them either asserted or negated. Since it is independent, we absolutely need it to complete the ranking requirements.

We close by noting that the test could have been simplified by an earlier finding: ERC [3ba] entails [1ca],[2ca], and [2cb]. Anything inconsistent with [3ba] is inconsistent with its consequences. The whole operation could have been conducted in terms of [3ca] and [1cb] alone. The final conclusion is that the entire content of the data under consideration boils down to just two independent ERCs.

(48) **The Delete-All Grammar**

ERC#	*CPLX	MWd	ONS	PHR	MAXC	*Remark*
[1cb]		**W**	**L**		**L**	pre-V **del** ≻ faith2
[3ba]	**W**			**L**	**L**	pre-P **del** ≻ faith

This is the 'Most Informative Basis' for the delete-all grammar, the most concise and (in a well-defined sense) informative representation of its structure. Interested readers should turn to Brasoveanu & Prince 2005 for an exploration of this notion and development of an algorithm that produces the Most Informative Basis for any set of ERCs.

6. The Meaning of ERC Entailment

What does it mean, in terms of data, to say $\alpha \Rightarrow \beta$, 'α entails β', for ERCs α and β? Since this supplies the basis for all arguments about implication between processes, it is useful to look inside this formally simple but moderately abstract expression.

An individual ERC deals with the relation between two candidates. It gives the ranking conditions under which one is guaranteed to be rated as better than the other.[36] In the C-deletion case, for example, ERC [3ba] tells us the ranking requirements needed to ensure this state of affairs:

⟨cost##→ cos.⟩ ≻ ⟨cost##→ cost.⟩
'it is better for phrase-final /cost/ to come out as *cos.* than as *cost.*"

Schematically: for violation-distinct candidates, the ERC [*p*≻*x*] precisely delimits the rankings under which candidate *p* will do better than *x*.

---

[36] I.e., if the two candidates are violation-wise distinct. If not, the resulting ERC is all *e*'s, and gives no information whatsoever about ranking. Optimality means being the best and, as the laws governing advertising famously recognize (explaining the proliferation of 'best' products), more than one thing can be best . All that's required is that nothing be strictly better. But only one *violation profile* can be optimal (Samek-Lodovici & Prince 2005.)

ERC entailment gives us a relation between two such delimiters of success: $\alpha \Rightarrow \beta$ means that whenever α is satisfied by a grammar, so also will β be. Suppose the ERC α looks at the comparison $[p \succ x]$ and the ERC β looks at $[q \succ y]$. The entailment relation, translated from constraint domination requirements into candidate relations, means this:

> ‘whenever *p* is better than *x* in some hierarchy,
> it must also be the case that *q* is better than *y* in that hierarchy.’

In the case at hand, we know that [3ca]⇒[2ca]. The meaning is

> “whenever it is better for **phrase-final** /cost/ to come out as *cos.* instead of *cost.*,
> it is also better for **pre-C** /cost/ to come out as *cos.* instead of *cost.*

Implicational universals are typically phrased along the lines of “whenever this process (input-output match) occurs, then so does that one.” This relates two optima: “whenever *p* is optimal, then *q* is also optimal.” But optimality is a derived notion, constructed from ‘better than’, and depends upon *p* and *q* each being *better than* lots of other things. It is these competitions, construed pairwise, that we have a direct grasp of. The place to look for implications is among them — among the ERCs they give rise to.

To say ‘candidate *p* is optimal’ is to say that *every* ERC of the form $[p \succ x]$ holds, for all *x* in *p*’s candidate set. Since the number of distinct ERCs is limited,[37] asserting optimality means asserting a finite set of ERCs. Each optimum gives rise to at least one ERC set which, when satisfied, ensures its optimality. (Typically, there will be a number of logically equivalent sets of this character.) Let’s call any such set of ERCs a *guarantor* for the candidate. The ERCs in a guarantor impose ranking conditions that are individually necessary and collectively sufficient for the optimality of the item under its guarantee.

Any standard implicational universal ‘optimal *p* ⇒ optimal *q*’ now becomes investigable as a statement about the guarantors for *p* and *q*. Namely: given any guarantor for *p*, we are assured of the validity of *all* the ERCs in any guarantor for *q*. We used exactly this kind of reasoning in our discussion of the *t,d*-deletion case. The guarantor for pre-pausal deletion consisted of a single ERC, [3ca], as did the guarantor for pre-pausal retention, ¬[3ca] = [3ac]. This made life easy.

For pre-C deletion, the data presented us with a two-ERC guarantor, but we argued our way down to one (one of the ERCs being entailed by the other). For the pre-V deletion guarantor, we demonstrated that the two ERCs in the data are logically independent and therefore both necessary. What we found was that one of them, [1ca], calling for *Cplx≫MaxC, was itself a guarantor for pre-C deletion, and via its negative [1ac] yielded MaxC≫*Cplx, sufficing for aligned retention in both pre-C and pre-pausal environments.

Once we see how entailments develop from ERC structure, it becomes clear that many other patterns must surely be waiting to be found. Just as there are entailments from sets that do not

---

[37] A generous upper bound is given by the total number of distinct nontrivial ERCs over n constraints, $3^n - 2^{n+1} + 1$. Nontrivial = contains both W and L. The first terms counts the number of ERCs over {W,L,e}, the second term subtracts off the number of ERCs over {W,e} and the number over {L,e}, the third term compensates for counting the degenerate ERC consisting of all *e*’s in the previous terms. This upper bound is generous because it does not attempt to assess the size of the largest *consistent* sets of nontrivial ERCs.

follow from a single ERC alone, so there will be implications from sets of optima to other (sets of) optima, and, concomitantly, patterns of mutual inconsistency involving several optima. To get a sense of how this can develop, consider an example:

(49) **Entailment from sets of Optima** .

	C_1	C_2	C_3
[1]	**W**	**L**	
[2]		**W**	**L**
[3]	**W**		**L**

Here we have [1] and [2] entailing [3], although neither succeeds individually.[38] If [1] and [2] are drawn from guarantors for distinct optima, and if [3] is the (entire) guarantor for yet another optimum, then we have a linguistic result: any grammar admitting the optima for [1] and [2] must also admit the optimum for [3]. This carries along with it a typological inconsistency result: ¬[3] cannot coexist with the other two. Further characteristic modes of entanglement between ranking conditions can be expected to emerge when expectations in the field advance to the point where logical and ecological situations are routinely scrutinized in more detail.[39]

To obtain a kind of base-level view of ERC entailment, it is instructive to pursue the matter all the way back to its grounding in violation patterns.

In considering whether an entailment relation $\alpha \Rightarrow \beta$ holds between two ERCs α and β, we must examine every constraint C to see whether $C[\alpha]$, the comparative value C assigns to α, stands in the appropriate relation to $C[\beta]$. Recall that the course of entailment runs L→e→W, further articulating the boolean pattern F→T.

There are three cases to consider, one for each value that C can assign in α.

(50) **Violation Patterns behind ERC entailment**

(i) If $C[\alpha] = L$, then $C[\beta]$ = L,e,W. The comparative value L like the truth value F entails anything (*ex falso quodlibet*). If C is forced to be subordinated in α, then an entailed β is locally the same or weaker in its requirements on C.

To spell this out at the level of violation patterns, we write $C(z)$ for the number of violations C assigns to any candidate z, and we take α to be the ERC $[p \succ x]$ and β to be $[q \succ y]$.

If C earns L in ERC α, which compares p to x, then the desired loser x is perversely doing better than the desired winner p. In violations, $C(p)>C(x)$.

So: for $\alpha \Rightarrow \beta$, **if $C(p)>C(x)$ then $C(q)$ and $C(y)$ are unrestricted**, allowing $C[\beta]$ to take on any comparative value.

(ii) If $C[\alpha] = W$, then we *must have* $C[\beta]=W$. If C is a possible dominator in α, then it must remain one in any entailed β, to ensure that every ranking under which α is true is also one in which β is true.

$C[\alpha] = W$ means that desired optimum p does better than it competitor x. In this case, when $\alpha \Rightarrow \beta$, it will also be the case that β's desired optimum q does better than its competitor y.

So: $\alpha \Rightarrow \beta$ requires that **if $C(p)<C(x)$, then $C(q)<C(y)$**.

This is the violation content of the fact that a W-antecedent demands a W-consequent

---

[38] Specifically, [1]∘[2]→ [3] by L-retraction.

[39] And indeed there may be significant value in fashioning an implication finder like Anttila & Andrus's that operates in the ERC context.

(iii) If C[α] = e, then it must be that either C[β] = e *or* C[β] =W. If C is absolutely unconstrained by α, then it cannot be constrained to be dominated in an ERC entailed by α, but it is free to participate as a possible dominator.

So: to get α⇒β, **if C(*p*)=C(*x*), then C(*q*)≤ C(*y*)**.

ERC theory provides the essential tools for exploring the structure of grammars and their relationships to each other. As a measure of its power, it is worth stepping back to appreciate the amount of condensation inherent in the notion of ERC entailment.

- A candidate relates two linguistic representations – input and output.
- A basic ERC derives from a comparison of two candidates.
- Simple entailment relates two ERCs.

An expression like 'α ⇒ β' is therefore talking about — minimally — some 8 linguistic representations. Fusion, which comprehends entailment and inconsistency in all of their generality, will bring in yet more. In working at this level, we control complex, multi-candidate patterns of inequalities in violation structure via a few simple rules of manipulation.

7. How ERC entailment generalizes Harmonic Bounding

Harmonic bounding arises from the order structure of candidate sets. So many candidates, so few optima! (For the inveterate loser, something always gets in the way.) In the simplest case, there is one identifiable blocker, one candidate that is never beaten by the harmonic boundee on any constraint and which beats it at least once. This guarantees that the unfortunate boundee can never be optimal, because we can identify another candidate (which itself need not be a possible optimum) that is always better, no matter what the ranking is.[40]

ERC-wise, we are looking at the comparison between a candidate futilely desired optimal, call it *z*, and a single more successful competitor, *q*. The ERC [*z*≻*q*] contains no W's and at least one L. It belongs to L^+. No ranking satisfies it. Turned the other way, its negative [*q*≻*z*], which contains only W's and *e*'s, imposes no conditions and holds under any ranking. Harmonic bounding emerges as the existence of an ERC that is everywhere valid (and whose negative is nowhere valid).

Entailment is a relation between two ERCs. When α⇒β, the first lays down conditions under which the second is sure to hold. In any grammar validating α, the requirements of β must also be met. Instead of being everywhere valid, without restriction, the ERC β is guaranteed valid in a precisely-defined subset of rankings. This generalizes the notion of harmonic bounding — unconditional validity — by allowing us to place conditions on the circumstances in which a statement like *q*≻*z* must be true.[41]

ERC entailment generalizes harmonic bounding in another fundamental respect. Bounding is a phenomenon that is limited to within-candidate-set relations. It makes no direct sense to compare candidates from different candidate sets; the candidate set determines the zone of competition and harmonic bounding is all about the picking of optima from that zone. But the ERC knows nothing of candidates and candidate sets: it is phrased entirely in terms of conditions

---

[40] This reduces the fact that the bounder beats the boundee on *n*! rankings to their local performance on *n* constraints.

[41] In α⇒β if the ERC α is valid, and therefore true of every ranking, no restriction is placed on the set of rankings in which β holds, so it must hold everywhere, and we're back to harmonic bounding.

on the ranking of constraints. We evaluate the truth of $\alpha \Rightarrow \beta$ quite independently of whether α and β have any particular substantive connection. Recall the core consequence of an entailment like $[p \succ x] \Rightarrow [q \succ y]$:

$$\text{if } p \succ x \text{ then } q \succ y.$$

This establishes a relation between the performance of p (wrt x) and the performance of q (wrt y), regardless of what p and q pertain to. There is no hint of restriction to the same inputs. Though derived from within-candidate-set information, the ERC is exportable in a way that harmonic bounding information is not.

It is precisely this independence of candidate set that allows ERC entailment to relate quite different mappings. In the Kiparsky-Anttila *t,d*-deletion example, the typological implications emerge from cross-candidate-set ERC relations. Because ERC entailment is a true *generalization* of bounding, including it as a special case, we also use it to illuminate candidate-set-internal structure as well, assessing the dependence and independence of ERCs arising from the competitions from a common input.

The same considerations apply to bounding in its most general form, when a candidate z is kept from optimality not by a single candidate, but by a gang of candidates that cooperate to ensure that z can never win. In the general sense, a candidate z is harmonically bounded when there is a set of candidates $Q=\{q_1,\ldots,q_n\}$ which has the property that Samek-Lodovici and Prince (1999:9, 2005:4) call 'reciprocity': whenever z is better than q_i on some constraint, there is a q_k that is better than z on that constraint, covering as it were for q_i. The set Q functions collectively to ensure that there is always something better than z on any ranking.

Crossing from the candidate-side to the constraint-side, via ERC talk, we have a collection of ERCs $[z \succ q_1],\ldots[z \succ q_n]$, such that no ranking can satisfy them all simultaneously. Whenever any of them earns W on a constraint (because for that constraint z is better than a competitor q_i), there is always another ERC that earns L, because it involves a q_k better than z. This shows that the fusion of the entire set contains no W's, since any column that contains a W also contains an L. It belongs to L^+, and is universally invalid.

Using the negative, we can turn this fusional ERC into one that is *valid* under all rankings.

$$\neg(\, [z \succ q_1] \circ \ldots \circ\ [z \succ q_n]\,)$$

As with the simple bounding case we began with, collective harmonic bounding reduces to the existence of a certain kind of (in)valid ERC.[42] The structure of the ERC displays the kind of generality noted above: a fusional composite is in no way restricted to ERCs from a single candidate set, and it seamlessly folds the notion of harmonic bounding into the broader notion of 'inconsistent ERC set'.

A more abstract but even closer relation between harmonic bounding and ERC entailment is disclosed when entailment is itself understood to involve *order*. Entailment is reflexive, symmetric, and transitive — a partial order on the objects that it relates. This comes out in many ways, but is directly reflected in propositional-type logics at the level of truth values. Suppose we impose an order on truth values of the form F<T. (This can be concretized by identifying F with 0 and T with 1, for example.) Then, for any assignment of truth values, the conjunction of two formulas takes on the *minimum* value exhibited in the conjuncts (i.e. false if any is false).

---

[42] There is also a corresponding entailment formulation. Suppose that $Q \cup \{z\}$ is minimal in the sense that removal of any of q_i yields a *consistent* set. (If Q is not minimal, we can make it so by tossing out elements until it is.) Then it can be shown that the coresponding ERCset $\{[z \succ Q \backslash q_k]\} = \{[z \succ q_i] \mid q_i \neq q_k\}$ entails $\neg[z \succ q_k] = [q_k \succ z]$ for all $q_k \in Q$.

Material implication, P⊃Q, is true iff the value of P is less-than-or-equal-to the value of Q; *i.e.* it is false only when |P|>|Q|, when the antecendent is true and the consequent false.

The logic of OT as disclosed in ERC theory is three-valued, but the same ideas apply *mutatis mutandis*.[43] Suppose we impose an order on the comparative values of this form: L<e<W.[44] We now have an order statement of ERC entailment that is identical to the prop calculus pattern. We may say 'α entails β' for distinct nontrivial ERCs, just in case for every constraint, the comparative value of α on that constraint is *less than or equal* to the comparative value of β.

This parallels perfectly the way harmonic bounding works on candidates. If, for non-identical candidates *q* and *z*, we have $|q|\leq|z|$ *in violations* on every constraint, then *q* harmonically bounds *z*. A simple harmonic bound is just a lower bound in the coordinate-wise order on the violation vectors. An entailer is just a lower bound in the coordinate-wise order on comparative vectors.[45]

In summary: ERC entailment generalizes harmonic bounding in two crucial respects. First, it allows us to restrict the success of a better-than-or-equal-to relation to a subset of rankings, where harmonic bounding insists that such a relation hold across the board. Second, because the ERC is completely exportable from the candidate set in which it arises, ERC entailment expresses relations between comparisons which need have no input in common, taking this general form:[46]

if $p\succ x$ then $q\succ y$

This places ERC entailment at the center of arguments about relations between optima.

8. Across the Great Divide

OT deals with two basic objects: constraint hierarchies and candidate sets. Statements about one are often mirrored in statements about the other. As is often the case in such situations, it may be easier or more illuminating to work with one way of characterizing an issue than with another, even when they are equivalent over the problem at hand. In OT, this means working on the *candidate-side*, with relations between candidates, or on the *constraint-side*, with relations between constraints. The ERC connects specific constraint-side better-than relations with specific constraint-side domination requirements.

---

[43] "This phrase and the use of it may best be explained by an example. A proprietor of an estate feus his lands, and the feu contracts all contain the same general clauses, the same obligations on the feuars and confer the same rights. In such a case two of the feu charters are said to be the same mutatis mutandis, that is, they are the same, if (or when) the name of the disponee, the particular description of the lands feued, and other such-like particulars which are peculiar to each, are changed." http://www.clickdocs.co.uk/glossary/mutatis-mutandis.htm.

[44] First introduced in Meyer 1975, discussed and explored in Prince 2002a and further discussed in Brasoveanu & Prince 2005. Observe that fusion is *not* minimal in this order — another order is required: L<W<e. The operation that is minimal in the L<e<W order is the three-valued analog of conjunction, as interpreted by Lukasiewicz and Kleene. See Prince 2002a:51ff.

[45] The two notions come together nicely when the ERC develops from within a single candidate set. Given a desired optimum *q*, we can identify each candidate with an ERC, candidate *x* with ERC $[x] = [q\succ x]$. Then if *x* bounds *y*, $[x]$ entails $[y]$. See Prince 2002a:41. The converse is not necessarily true: see below for an example.

[46] Recall that *p* and *x* are of the form $\langle a\rightarrow k_1\rangle$ and $\langle a\rightarrow k_2\rangle$, while *q* and *y* will be of the form $\langle b\rightarrow u_1\rangle$, $\langle b\rightarrow u_2\rangle$.

As a first example, consider the unfolding of typologies. With *n* constraints, we famously have *n*! formal hierarchies (linear orderings of the constraint set), the fact that has historically caused such distress to those basking in the low-wattage glow of their own innumeracy. Using exhaustive search, we might seek the distinct grammatical systems (patterns of input-output relations) with admirable but uninspired thoroughness by cranking through the hierarchies one at a time. This is the constraint-side approach. On the candidate side, one might observe that there's a limited number of candidate sets worth considering, and in each of them a limited number of serious candidates to consider. The number of possible *candidate* combinations — pick one candidate from the first set, one from the second, and so on — could well be very much smaller than the number of hierarchies. Yet the distinct grammars are exactly those with differing patterns of optima! Recall, for example, the data under consideration by Anttila and Kiparsky, repeated here for convenience:

(51) **Violation Table for the Constraint Set**

IN	OUT	*Complex	MWd-σ-init	Ons	Ph-MWd-fin	MaxC	*Remark*
1. cost us	a. cost.us	*		*			faith1
	b. cos.tus		*				faith2
	c. cos.us			*		*	***del***
2. cost me	a. cost.me	*					faith1
	b. cos.tme	*	*				faith2
	c. cos.me					*	***del***
3. cost##	a. cost.	*					faith
	b. cos.				*	*	***del***

Here we have 5 constraints, yielding 5!=120 hierarchies. But the number of distinct candidate combinations is merely 3×3×2 =18. Some preliminary scouting of the violation table would reveal what we noticed above, that candidate ⟨2b⟩ is harmonically bounded by ⟨2a⟩, and can be dispensed with.[47] This reduces the game to 3×2×2= 12, a mere tenth of the constraint-side labor, and a veritable invitation to exhaustive search.[48] In fact, optima-calculating programs tend to work on the candidate-side.

Harmonic bounding provides an example of a central structural problem that appears in importantly different guises depending on how it is viewed. From the candidate-side, it is an observation about the ordering structure within a candidate set: such-and-such can never be optimal because we can designate a collection of candidates that always get in the way, as described by the Reciprocity Condition. On the constraint-side, harmonic bounding amounts to unsatisfiable ranking requirements. In the case of (51), we see (candidate-side) that candidate ⟨2b⟩ is always worse than candidate ⟨2a⟩, whatever the ranking. The constraint-side situation associated with [2b≻2a] is that the constraint MWd-σ-init must be dominated — but there is no possible dominator.[49] The information on display on each side is quite different in character.

This gives us two distinct ways to explore many kinds of issues. In the case of harmonic bounding, we can, with Samek-Lodovici & Prince 1999, assume that the Reciprocity Condition is met, and following a chain of *iff*'s, show that no ranking exists. Or we can assume that no

---

[47] The harmonic bounding relation is detected as ERC entailment in the analysis given above. See Prince 2002a:35ff, where the generality of this effect is demonstrated.

[48] The number will shrink further if we make use of the constraint-side implications discussed above.

[49] The ERC is invalid, but, as remarked above in fn. 33, p. 22, it contains more information than a bald assertion of nonexistence of a ranking. It identifies the exact source of the problem.

ranking exists, and derive Reciprocity. Progress along the later route is aided considerably by the development of ERC theory. We know now that nonexistence of ranking is co-extensive with the existence of a set of ERCs that fuse to L^+. An inquiry into the meaning of this for candidate-set structure will unearth the Reciprocity phenomenon on the candidate-side. Both modes of analysis are valuable and necessary; they may open different perspectives, reveal different kinds of patterns, be relatively swift or relatively labor-intensive in arriving at a desired result, generalize or fail to generalize in desirable ways. It is incumbent upon the analyst to use all tools available — and to construct sharper new ones when possible.

Anttila approaches implicational problem in purely candidate-side terms. This may be entirely sensible in many circumstances, but we have seen that venturing across the bridge provided by the ERC to the constraint-side can have a significant pay-off.

9. Back to the Future

Where are you coming from? Where are you going?
-Plains greeting (*via* Hu Matthews, 1971)

The 'Technique of Necessary and Sufficient Conditions' is identified in OTCIGG:222 as a method for establishing implicational universals.

> Step 1. Determine the necessary and sufficient conditions on the ranking of constraints in a hierarchy in order that each of the relevant structures be admitted into the inventory by that constraint ranking.
>
> Step 2. Examine the logical entailments that hold among these conditions: arguments of the form: in order to admit structure φ it is necessary that the constraints be ranked in such-and-such a way, and this entails that the constraint ranking meets the sufficient conditions to admit structure ψ.

This is stated in terms of output inventory structure, but applies without modification if we read 'structures' as entire candidates (input-output pairs). It is the method we used above to derive implications from Kiparsky's constraint system. Basically, we start on the candidate-side and move to the constraint-side to establish implications in term of the logic of constraint domination. We then move back to the constraint-side to redeem constraint entailments for candidates. ERC theory allows us to pursue the method in terms of the narrowly-specificiable intrinsic logic of OT, rather than pitching ourselves headlong into the wilds of the propositional calculus.

9.1 Harmonic Completeness of POA

To gain a broader feel for the usefulness of the tools developed above, let us re-work some of the argument patterns found in OTCIGG. We will examine three instructive cases. We consider first a typical markedness-type implication 'complex ⇒ simple', set in the realm of Place of Articulation (POA). The thesis to be demonstrated is this: that the presence of complex consonants with secondary articulations entails the presence of simple consonants with each of the places of articulation that combine in the complex case (OTCIGG:223). The background theory of POA is a simplified one, based on Clements 1991, and making use of pre-correspondence mechanisms (which will be characterized slightly differently from the original.)

The place node may be specified for none, one, or two features; specifications are marked as primary and secondary. Features may be removed, parsed as primary or secondary (regardless of their original demarcation); but, in accord with the 'containment' view, they may not be changed or inserted. Accordingly, the relevant constraints are (slightly rephrased):

Markedness:

*PL/φ	The PL node does not contain the feature φ.
*PL/$\langle\ \rangle$	The PL node must dominate some feature.

Faithfulness:

F:PARSE-feat	Every feature specification in the input must appear in the output
F:$2\nrightarrow 1$	An input secondary articulation feature does not emerge as a primary articulation.

To show that 'complex⇒simple', let us consider the treatment of the generic complex input place node, $\langle\pi{:}1\ \psi{:}2\rangle$. To concretize the discussion, without loss of generality, let's pick π = lab and ψ=cor and consider $\langle$lab:1 cor:2$\rangle$, the place node for a palatalized labial like p^j.

We first observe that if $\langle$lab:1 cor:2$\rangle$ is in the *output*, it can only come from itself, as it were; it must be the result of faithful mapping. Since features cannot be inserted, both must be present underlyingly. The only other input choice for it would be $\langle$ cor:1 lab:2$\rangle$, with opposite assignment of features to primary and secondary places, designating a labialized coronal like t^w. But flipping of primary and secondary POA is harmonically bounded by simple faithful replication. Faithful reproduction of /t^w/ as t^w bounds flipping it to p^j, so /t^w/ cannot serve as the input for surface p^j. In the general notation, identity-mapping of $\langle\psi{:}1\ \pi{:}2\rangle$ bounds flipping it to $\langle\pi{:}1\ \psi{:}2\rangle$, so it cannot underly $\langle\pi{:}1\ \psi{:}2\rangle$.

To establish the claim, we need only consider placeless alternatives. In the case of underlying simple articulations $\langle$lab$\rangle$ and $\langle$cor$\rangle$, the *only* possible competitor to either one is the placeless output $\langle\ \rangle$, since features cannot be inserted or changed. It follows that besting the placeless competitor is sufficient for the optimality of faithful mono-specified forms. Besting the placeless is *not* sufficient to guarantee that the bi-specified form $\langle$lab:1 cor:2$\rangle$ to be self-realized (further candidates must be defeated), but then it is only necessary that it be necessary. That is: to show that complex⇒simple, we *assume* a grammar which has complex $\langle$lab:1 cor:2$\rangle$ among its outputs, and pursue the consequences of *any* ERC that is necessary to achieve this. Among them will be the one which declares how faithful $\langle$lab:1 cor:2$\rangle$ bests the placeless realization $\langle\ \rangle$.

Let us therefore examine each of the relevant inputs, running faithful reproduction against the placeless output. We present the result as a comparatively-annotated tableau. The constraints are *not* listed in ranking order — ranking is what's under question. Since this will be true for all cases we discuss, we abandon the dashed-vertical-line notation for nonranking.

(52) **Complex ⇒ Simple**

	*PL/lab	*PL/cor	*PL/⟨ ⟩	F:PARSE-Feat	F:2↛1
⟨lab:1 cor:2⟩ → ⟨lab:1 cor:2⟩	*	*			
↛ ⟨ ⟩	**L**	**L**	* **W**	** **W**	
⟨lab⟩ → ⟨lab⟩	*				
↛ ⟨ ⟩	**L**		* **W**	* **W**	
⟨cor⟩ → ⟨cor⟩		*			
↛ ⟨ ⟩		**L**	* **W**	* **W**	

It is mere eyeball work to determine that the last two (sufficient) conditions for the optimality of the simple articulations are entailed via L-retraction by the topmost ERC, which is generated from the victory of the complexly-specified form over its placeless competitor. QED.[50]

9.2 Possible Onsets in Sonority-based Syllable Theory

A similar facilitation occurs with the kind of arguments developed in the 'universal syllable theory' of OTCIGG: Ch. 8. The goal there is to transcend the basic syllable theory (Ch. 6), which assumes a prior division of phonological elements into one margin-worthy and one peak-worthy segment type — notated C and V respectively — following Clements & Keyser 1979. The generalized theory goes from C *vs.*V to a scalar spread of sonority types spanning the dichotomy, and allows for a range of sonority-based structural restrictions. The goal is to show that scalarization leads to a rich set of inventory implications, none stipulated or even anywhere close to stipulable in the theory, whose empirical consequences if not an exact transcription of observed or conjectured reality at least bear a discernible relation to it. An example is the 'onset/coda licensing asymmetry', restated from OTICGG:154, ex. (194):[51]

(53) **Possible coda segment ⇒ Possible onset segment.** The inventory of possible codas in any language is a *subset* of the inventory of possible onsets in that language.

This is observationally falsified, for example, by the English language and its name (ŋ is not a possible onset), but remains reasonable as a hypothesis about a core pattern impinged upon by

---

[50] The OTCIGG:223 argument compares its equivalent of ⟨lab:1 cor:2⟩ to a competitor ⟨lab:1 ~~cor:2~~⟩, showing that ⟨cor⟩ will then be among the outputs, noting that a "similar but slightly more complex argument" guarantees the presence of ⟨lab⟩. ERC theory flattens out the complexity of such arguments. In this case, handling the disjunctivity of *PL/⟨ ⟩ and F:PARSE-Feat is not much more of a strain than if there were only one of them to consider.

[51] Observe that the term 'licensing' is a metaphor for the emergent results of the entire system rather than a description of how its individual constraints work. There is no licensing in OT. A licensing condition is one, that when met, ensures goodness. But satisfying a constraint, or even better-satisfying one, is no guarantee that a form will be admitted or that the constraint's demands will be met.

various other independent factors.[52] OTCIGG studies how it devolves from a set of assumptions about sonority and relative suitability for peak or margin position , assumptions which make no mention at all of inventories or possibility. This leads not only to a substantive linguistic theory, but also provides a model of how scalar phenomena — rife in language — can be made sense of.

In the present context, it is illuminating to reconstruct at least part of the argument, which is necessarily built from a number of results.[53] First, the background. Having replaced the two-way C/V classification with a range of sonority-defined elements, we posit sets of constraints that discourage the affiliation of each such segment-type with the canonical C- and V-positions, the syllable Margin (onset, coda) and Peak (nucleus). The constraints are arranged in two fixed hierarchies, based on sonority, echoing the hypothesis that Peaks favor higher-sonority, more prominent elements, while Margins are exactly opposite in their segmental affinity.

(54) **Peak Hierarchy** (fixed). *P/t ≫ …*P/u ≫*P/e≫ *P/a

In the context of the theory, this will tell us, *e.g.*, that *a* is a better syllable nucleus, a better V as it were, than *t*; that *e*, being more sonorous, is better as a nucleus than *u*, and so on.

(55) **Margin Hierarchy** (fixed). *M/a≫*M/e ≫*M/u≫ … ≫ *M/t

This incorporates the intelligence that *t* is a better syllable margin, a better C, than *a*,[54] and draws a number of finer-grained distinctions among other segments based on relative sonority. The hierarchies are derived from primitive scales of intrinsic prominence that are not themselves preferential (OTCIGG:161-162).

These hierarchies are fixed in the sense that their domination relations must be the same in all grammars, constrasting with the otherwise free permutation of constraints.[55] In present terms, they correspond to a set of ERCs deemed present in every grammar, without requiring candidate sponsorship. (Since an ERC is at heart a statement about constraint domination, it can be used to express domination relations directly, whether or not it is derived from data.)

The other markedness constraints are familiar from the basic syllable theory:

---

[52] It is clear observationally that codas typically admit *fewer* elements than onsets and never admit more; this can happen without inclusion of one inventory in the other. Since the claimed Prague-like subsetting relation is not a summary of observations, it will have to be justified inferentially, by its role in a more embracing theory.

[53] The structure of the argument, as compelled by the structure of the theory, is laid out in OTCIGG:154-155.

[54] Observe that *a* is not entirely hopeless as a margin: the 'velar glide' of Axininca Campa appears to be a marginal (onset) *a*. SeeYip 1983, McCarthy & Prince 1993:155-159 for discussion.

[55] The alternative to fixed hierarchies like *a≫*b≫*c is to define the constraints in a 'stringency hierarchy' of inclusiveness of the form *{a}, *{a,b}, *{a,b,c}, in which each constraint in the hierarchy rejects a superset of its predecessors rejectees. This preserves the underlying scalar structure and allows for free ranking while generating the possibility of further interactions. Work along these lines includes Green 1993, Kiparsky 1993, Prince 1997-2001, de Lacy 2004, 2006. On the related issue of 'elsewhere effects', see Bakovic 2006.

(56) **Syllable structure markedness** constraints

a. ONS. Syllables must have an onset node.
b. NOCODA. Syllables must not have a Coda node.

The faithfulness system in OTCIGG is somewhat different from that often assumed in the wake of McCarthy & Prince 1995: under it, Gen can neither change nor insert feature values. For ease of comprehension, we update it in the direction of correspondence theory, but not so much as to alter the original argument. Let's assume with OTCIGG that syllables crucially have a hierarchical constituent structure, in which σ dominates margin nodes Ons and Coda and the peak node Nuc, where only the Nuc node is obligatory. Assume that all output strings are completely syllabified. Segments can be deleted, but we will assume for present purposes that only *featureless segments* can be inserted (that is, lack input correspondents). As in OTCIGG, we'll assume that a phonetic component post-processes the output of phonology and fills in feature values; this means that the actual values of inserted segments are not visible to phonology (a clear shortcoming which is transcended in full-blown correspondence-based faithfulness theory). This gives us three relevant faithfulness constraints, stated informally here, which can be formulated in the usual correspondence terms:

(57) **Faithfulness Constraints** .

a. F:MAX. Segments may not be deleted.
b. F:DEPC. No inserted (empty) segment occupies a margin node (Ons, Coda).
c. F:DEPV. No inserted (empty) segment occupies the peak (Nuc).[56]

Since the goal of the enterprise is to determine the relation between *possible* onsets and *possib*le codas, one of the basic preliminary questions will have to be this one:

> What ranking conditions are required to ensure that a given segment is a *possible onset* in some grammar?

To investigate, let us proceed concretely by determining when an input of the form /ue/ is realized with /u/ as an onset and /e/ as a nucleus — in short, as [we]. For emphasis, we will write the marginalized version of [u] as 'w' and the marginal *e*, a glide, as *ĕ*, but we are assuming featural identity and only positional difference.

The conclusions we aim to draw are not limited to 'u ' and 'e'.What's important to get the analysis rolling is that MAX≫*M/w, rather than vice versa, so that *u* can't just be deleted outright to satisfy *M/w, which declares 'no *u* as onset or coda'. With this ranking, *u* (spelled *w*) is said to be a *tenable margin*. Similarly, we want MAX≫*P/e, so that *e* won't be deleted to satisfy *P/e, 'no *e* as nucleus'. Such 'e' we will call a *tenable peak*. We seek to ensure that tenable margins and peaks are actually realized, somewhere, as real margins and real peaks (respectively).

Proceeding systematically, we gather up all competitors that we might need to establish the optimality of ⟨ue→.we.⟩. The tabulation is slightly modified from OTCIGG:173 ex. (237), though all candidates are retained.[57]

---

[56] Under the view of Gen taken here, DEPC and DEPV could be regarded as markedness constraints banning the empty segment from the surface. To emphasize their essential continuity with current assumptions, we classify them as faithfulness due to their effect on correspondence. All versions of DEP overlap with markedness, since insertion involves the presence of an output element whose qualities are up for judgment.

(58) **Peak vs Margin Parsing Competition**

	/ue /	Ons	*P/u	Depc	DepV	*M/ ĕ	*M/w	*P/e	NoCoda
0	.we.						*	*	
1	.cu.ce.		*	**				*	
2	.u.e.	**	*					*	
3	.u.ce.	*	*	*				*	
4	.cu.e.	*	*	*				*	
5	.wV.ĕV.				**	*	*		
6	.wVĕ.				*	*	*		*
7	.uĕ.	*	*			*			*
8	.u. ĕV.	*	*		*	*			
9	.cu.ĕV.		*	*	*	*			
10	.wV.e	*			*		*	*	
11	.wV.ce.			*	*		*	*	

We use the following notation to iconize the goings-on. Marginal elements are typeset small, peaks large. Inserted (empty) segments are symbolized c when in margin and V when in peak. All deletional competitors are excluded, *ex hypothesi*, since they will lose on Max, under the assumption of margin-tenability for *w* and peak-tenability for *e*. An infinity of harmonically-bounded epenthetic candidates has been omitted, as licensed by the Fill Violation Theorem of OTCIGG:118, which enumerates positions (a mere handful) into which epenthesis can emerge as optimal under the constraint interactions of the syllable theory.

Extracting sense from the display might seem to require mastery of an art akin to haruspicy or the interpretation of fly-specks. The reason for this impression is the same as before in our dealings with t,d-deletion: the relevant information has not been calculated. Once the comparative structure is laid out, it will quickly become apparent that only two competitors are needed to determine onsetting of /u/ — ⟨1⟩ and ⟨2⟩ — and only one more — ⟨5⟩ — to secure peak position for /e/.

Here is the comparatively annotated version of the same array, divided into bands where the crucial action takes place. The desired optimum occupies row 0 .

---

[57] Changes: constraint names as noted; phonetic targets; notation; and the rows are rearranged according to the following scheme: 0:a, 1:h, 2:d, 3:e, 4:g, 5:l, 6:c, 8:f, 7:b, 9:i, 10:j, 11:k.

(59) **Comparatively annotated tableau**

	/ue /	Ons	*P/u	Depᴄ	Depᴠ	*M/ ĕ	*M/w	*P/e	NoCoda
0	.we.						⊛	⊛	
1	.ᴄu.ᴄe.		* **W**	** **W**			**L**	*	
2	.u.e.	** **W**	* **W**				**L**	*	
3	.u.ᴄe.	* **W**	* **W**	* **W**			**L**	*	
4	.ᴄu.e.	* **W**	* **W**	* **W**			**L**	*	
5	.wᴠ.ĕᴠ.				** **W**	* **W**	*	**L**	
6	.wᴠĕ.				* **W**	* **W**	*	**L**	* **W**
7	.uĕ.	* **W**	* **W**			* **W**	**L**	**L**	* **W**
8	.u.ĕᴠ.	* **W**	* **W**		* **W**	* **W**	**L**	**L**	
9	.ᴄu.ĕᴠ.		* **W**	* **W**	* **W**	* **W**	**L**	**L**	
10	.wᴠ.e	* **W**			* **W**		*	*	
11	.wᴠ.ᴄe.			* **W**	* **W**		*	*	

[SPOILER ALERT. The reader may wish to undertake the analysis before continuing.]

A first observation: the last two rows contain no L's, indicating that candidates ⟨10⟩ and ⟨11⟩ are harmonically bounded by the desired optimum, and require no ranking restrictions to fail. They can be ignored. Candidate ⟨11⟩ involves two adjacent epentheses to achieve a syllable structure no better than that of the entirely faithful desired optimum ⟨0⟩. (It is in fact excluded by the Fill Violation Theorem for that very reason.) Candidate ⟨10⟩ could be ruled out by an improved version of the theorem, since it uses epenthesis to obtain a worse syllable structure than that of the desired optimum.

Candidate ⟨2⟩, in failing, will take candidates ⟨3⟩ and ⟨4⟩ down with it. ERC-wise, we observe that [0≻2] entails, by W-extension, both [0≻3] and [0≻4]. We repeat the relevant rows in pure comparative format to demonstrate this.

(60) **Entailment #1**. [0≻2] ⇒ [0≻3], [0≻4]

/ue/ → .we.		Ons	*P/u	Depᴄ	Depᴠ	*M/ ĕ	*M/w	*P/e	NoCoda
[0≻2]	.u.e.	**W**	**W**				**L**		
[0≻3]	.u.ᴄe.	**W**	**W**	**W**			**L**		
[0≻4]	.ᴄu.e.	**W**	**W**	**W**			**L**		

The three losing candidates involved in these ERCs share the nucleus *u* and the presence of (at least one) onsetless syllable, flaws that the desired optimum lacks; and they best the optimum in only one respect, their lack of marginal *w*. In addition to these commonalities, candidates ⟨3⟩ and ⟨4⟩ suffer another flaw that the optimum lacks: epenthesis of C. This yields the entailment.

In sum: Candidate ⟨2⟩ must fail because of its shortcomings with respect to the desired optimum (on Ons, *P/u), and these are also shortcomings of candidates ⟨3⟩ and ⟨4⟩, which sport no other virtues than candidate ⟨2⟩.

Observe that harmonic bounding by candidate ⟨2⟩ is not at issue here. Because candidate ⟨2⟩ does *less* well on Ons than ⟨3⟩ and ⟨4⟩, it does not bound them (a bounder can never fail). To show this explicitly, we run ⟨2⟩ against the shared profile of ⟨3⟩ and ⟨4⟩. [58]

(61) Candidate ⟨2⟩ does not bound ⟨3⟩ and ⟨4⟩

	Ons	*P/u	Dep**C**	Dep**V**	*M/ ĕ	*M/w	*P/e	NoCoda
2	⊛⊛	⊛					⊛	
[2≻ 3=4]	* **L**	*	* **W**				*	

But this is irrelevant to ranking, since all do worse on Ons (to one degree or another) than the desired optimum, which doesn't violate it at all. ERC entailment has revealed a kind of relation, important for ranking, that harmonic bounding does not detect because it is concerned with *every* ranking, not just those meeting some prior condition (in this case, that candidate ⟨2⟩ loses to the desired optimum). There is a kind of order structure in the candidate set, relativized to a desired optimum, revealed by ERC entailment, which determines the *informativeness* of various suboptimal candidates. Candidate ⟨2⟩, though it is not better than ⟨3⟩ and ⟨4⟩ on all rankings, is more informative than they are with respect to the choice of ⟨0⟩ as optimum.[59]

It is also apparent that the ERC [0≻5] entails [0≻6] by W-extension.

(62) **Entailment #2**

/ue/ → .we.		Ons	*P/u	Dep**C**	Dep**V**	*M/ ĕ	*M/w	*P/e	NoCoda
[0≻5]	.w**V**.ĕ**V**.				**W**	**W**		**L**	
[0≻6]	.w**V**ĕ.				**W**	**W**		**L**	**W**

In both candidates, there is more epenthesis than in the desired optimum, which has none. Both marginalize *e*, which the optimum avoids. On top of these shared shortcomings, candidate ⟨6⟩ throws in a worse performance on NoCoda, meaning that whenever ⟨5⟩ is rejected in favor of the optimum, ⟨6⟩ will also be rejected.

Further entailments follow from the peak and margin hierarchies, which include the fixed relations *P/u≫*P/e and *M/ĕ≫*M/w. (Glosses: Low sonority *u* is a worse peak than higher sonority *e*; High sonority glide *ĕ* is worse marginally than low-sonority glide *w*.)

---

[58] Remark: we are not thinking of ⟨2⟩ as the desired optimum for the whole candidate set — we just want to see how it fares against some of its suboptimal confreres

[59] On informativeness, see Brasoveanu & Prince 2005.

(63) **PH MH and the like**

/ue / → .we.		ONS	*P/u	DEPc	DEPV	*M/ĕ	*M/w	*P/e	NOCODA
	PH		W					L	
	MH					W	L		
	PH∘MH		***W***			***W***	***L***	***L***	
[0≻7]	.we.≻ .uĕ.	**W**	**W**			**W**	**L**	**L**	**W**
[0≻8]	.we.≻ u.ĕV.	**W**	**W**		**W**	**W**	**L**	**L**	
[0≻9]	.we.≻cu.ĕV.		**W**	**W**	**W**	**W**	**L**	**L**	

Since *w* is indelibly a better margin than *ĕ*, and *e* irrefragably better peakwise than *u*, the desired optimum (*.we.*) puts its segmental content to better use than any of the competitors. As tableau (63) shows, the constraint logic requires consulting both hierarchies. Fusion of the two relevant hierarchy-defining ERCs, labeled PH and MH, does the trick, yielding the ERC PH∘MH which entails all three data-derived ERCs by various applications of W-extension.

Interpretive remark. All the suboptima here differ from the optimum along both peak and margin dimensions. Nucleizing the *u* in the suboptima avoids marginalizing it as *w*. The peak hierarchy PH helpfully sees nuclear *u* as a relatively bad idea (see the W's in the *P/u column). But it does not alone entail the ERCs based on candidates ⟨7⟩ - ⟨9⟩, because each suboptimum gains an additional apparent virtue, *lack* of marginal *w*, a 'flaw' conspicuous in the optimum *.we.* (Note the L's under *M/w, advertising loser success). The margin hierarchy is needed to dismiss this loser virtue as an illusion, *via* *M/ ĕ, and we're done.[60]

The number of W's flying around in tableau (63) indicates a messy logical situation, rich in disjunctions. For example, ERC [0≻9] reads:

$$(\text{*P/u}\gg\text{*M/w} \wedge \text{*P/u}\gg\text{*P/e}) \vee (\text{DEPc}\gg\text{*M/w} \wedge \text{DEPc}\gg\text{*P/e})$$
$$\vee (\text{DEPV}\gg\text{*M/w} \wedge \text{DEPV}\gg\text{*P/e}) \vee (\text{*M/ĕ} \gg\text{*M/w} \wedge \text{*M/ ĕ} \gg\text{*P/e})$$

For the ERC calculus, however, the entailment deduction provides little more difficulty than the very simple examples we began with.

Putting these observations together, we see that only three comparisons are needed to compel /ue/ to come out as *.we.* (Results shown as comparatively-annotated tableau without violations.)

---

[60] The ERC derived from ⟨8⟩ is also entailed jointly by those derived from ⟨2⟩ and ⟨5⟩, as shown in the following:

IIG	ONS	*P/u	DEPc	DEPv	*M/ĕ	*M/w	*P/e	NOCODA
0≻2	**W**	**W**				**L**		
0≻5				**W**	**W**		**L**	
fu	***W***	***W***		***W***	***W***	***L***	***L***	
0≻8	**W**	**W**		**W**	**W**	**L**	**L**	

(64) **Nonredundant ERC Set**

	/ue /	Ons	*P/u	DepC	DepV	*M/ĕ	*M/w	*P/e	NoCoda
0	.we.								
1	.Cu.Ce.		**W**	**W**			**L**		
2	.u.e.	**W**	**W**				**L**		
5	.wV.ĕV.				**W**	**W**		**L**	

All rows arel independent — none entails any other, nor does any pair entail the remaining third. Moving from the constraint-side to the candidate-side, we observe that only ⟨1⟩ and ⟨2⟩ are relevant to the onset issue. Like the desired optimum, candidate ⟨5⟩ realizes *u* as an onset, and were it to win, the conclusion would still stand that *u* qua *w* is a possible onset. We have therefore arrived at sufficient conditions for a *tenable* margin in any language to be a *possible* onset in that language, *i.e.* a real onset in some licit form.

(65) **Possible Onset Condition** (POC)

If α is a tenable margin, it is a possible onset iff the following two conditions *both* hold:

a. *P/α ≫ *M/α *or* DepC ≫ *M/α
b. *P/α ≫ *M/α *or* Ons ≫ *M/α

These conditions merely transcribe the ERCs [0≻1] and [0≻2]. (We return below to the question of necessity, as opposed to sufficiency, of these conditions.) In tableau form, they look like this:

(66) **POC** for α, a tenable margin

	Ons	DepC	*P/α	*M/α
(65)a		**W**	**W**	**L**
(65)b	**W**		**W**	**L**

Encouragingly, nothing more that 'α is better as a margin than as a peak' is needed to ensure that α will in fact show up as a margin — namely, as an onset. This argument plays out in ERC form as follows:

(67) **When α is a better margin than peak**

	Ons	DepC	*P/α	***M/α**
α better as **margin**			**W**	**L**
(65)a		**W**	**W**	**L**
(65)b	**W**		**W**	**L**

The first row, which announces that *P/α≫*M/α, clearly entails the other two, thus ensuring that when it's satisfied, both of their requirements will also be met.

When this condition fails, so that in some grammar (due to its own interleaving of the peak and margin hierarchies) a segment α is rated as more suitable for peak than margin, the POC tells us that it can still be coerced into marginhood by the action of the basic syllable structure constraints. When an onset is demanded (via Ons) and epenthesis is not welcome (via DepC), sonority-based considerations are overruled. The ranking conditions look like this:

(68) **When α is a better peak than margin**

	Ons	Dep**c**	***P/α**	*M/α
[1]: α better as ***peak***			**L**	**W**
(65)a		**W**	**W**	**L**
(65)b	**W**		**W**	**L**

The second and third rows are individually disjunctive, but the first row, taken with them, eliminates all disjunctions. To see the true situation more clearly, we fuse the first row with each of the other two, while retaining the first intact.[61]

(69) **A more informative rendition of (68)**

	Ons	Dep**c**	*P/α	*M/α
[1]: α better as *peak*			**L**	**W**
(65)a∘[1]		**W**	**L**	**L**
(65)b∘[1]	**W**		**L**	**L**

Conclusion: even when α emerges by ranking as a better peak than margin (row one), then if Ons and DepC dominate *both* of the relevant peak and margin conditions (second and third rows), α will still be a possible onset. As promised, this means that constraints of the basic theory (Ons, DepC) can join together to overpower sonority sensitivity.

This ERC-based discussion translates into a logically-equivalent restatement of the POC :

(70) **Possible Onset Condition**, restated.

If α is a tenable margin, it is a possible onset iff one of the following two conditions hold:

a. *P/α ≫ *M/α "better margin than peak"

b. Ons and DepC both dominate *M/α . "basic syllabic constraints prevail"

(OTCIGG:171, ex. (231))

By examining a certain input, we have determined ranking conditions that will guarantee that at least one output contains an onset of the desired type. Although we worked from a concrete example, the conclusions are general and apply to any segments with the right sonority relations. This does not, however, show that such conditions are *necessary* — that they accompany every onset. Could there be some other circumstances, undiscussed, in which the same kind of thing would happen — an underlying segment optimally parses as an onset — but under quite different rankings that do not include these?

To see that what we have examined is truly generic, providing conditions both sufficient and necessary, observe that the key losing candidates (and their associated ERCs) must appear whenever an onset shows up. Suppose, to be harmlessly concrete, that a certain *u* is an onset in some optimal output (and so for us written *w*). By Gen, it must be followed by a nuclear segment, call it α. There is a surely a suboptimal competitor in which the /u/ is nuclear, and onsets are supplied by epenthesis: ….**c**u.**c**α…. And this competes against the winner which has

---

[61] In so doing, we retain all global information while improving the local information content of latter two rows. Observe that no new W's are introduced, nor any L's taken away. Fusion is always guaranteed to produce a result jointly entailed by the fusands; in this case, entailment also runs in the other direction, in the sense that the fusional rows also entail the rows they replace. See Brasoveanu & Prince 2005 for further detailed discussion.

onset *w* and nuclear α, and in which every other structure is the same, thereby generating exactly the ERC [0≻1] of (59), which becomes POC (65)a. Similarly, there must be a competitor of the form …u.α… which differs from the winner *only* in that *u* and α are *both* nuclear, regenerating the image of candidate 〈2〉 and the ERC that goes along with it, which becomes POC (65)b. QED.[62]

9.3 Harmonic Completeness of Possible Onsets

These results also support a conclusion about the *harmonic completeness* of the proposed constraint system. Harmonic completeness along a certain structural dimension means that if a structure of a certain degree of markedness is admitted, then so are all structures less marked along that dimension.

(71) **Harmonic Completeness of Onsets with respect to Sonority**.

If τ is *less* sonorous than λ, then if λ is a possible onset in some language, so is τ.

OTCIGG:164, (215)

From the peak and margin hierarchies, the sonority assumption gives us *P/τ≫*P/λ (“τ is a worse peak than λ”) and *M/λ≫*M/τ (“τ is a better margin than λ”). Let’s put these in with the POC and see what happens. Recall the two necessary and sufficient conditions for possible onsethood of arbitrary α, from (66):

(72) **Possible Onset Condition for arbitrary α** (a tenable margin)

		ONS	DEP⊏	*P/α	*M/α
(65)a	POC1(α)		**W**	**W**	**L**
(65)b	POC2(α)	**W**		**W**	**L**

Since by assumption λ is a possible onset, we’ll instantiate the POC for λ, drop in the relevant pieces of the peak and margin hierarchies, and aim to conclude that possible onset conditions for τ are also satisfied.

(73) **POC + Peak and Margin Hierarchies**

	ONS	DEP⊏	*P/τ	***P/λ**	***M/λ**	*M/τ
PH for |λ|>|τ|			**W**	**L**		
MH for |λ|>|τ|					**W**	**L**
POC1(λ)		**W**		**W**	**L**	
POC2(λ)	**W**			**W**	**L**	

Fusing the MH and the PH with the POC conditions delivers the result:

---

[62] This argument reproduces that of OTCIGG:172.

(74) **POC, PH, MH fused**

	ONS	DEPC	*P/τ	*P/λ	*M/λ	*M/τ
PH			***W***	***L***		
MH					***W***	***L***
POC1(λ) ∘ MH ∘ PH		**W**	***W***	***L***	**L**	***L***
POC2(λ) ∘ MH ∘ PH	**W**		***W***	***L***	**L**	***L***

Italicized values in the last two rows are those introduced by fusion. These fused rows, by L-retraction, give us the satisfaction we seek, the POC for τ, as seen here:

(75) **POC satisfied for τ**

	ONS	DEPC	***P/τ**	*P/λ	*M/λ	***M/τ**
from POC1(λ)∘MH∘PH		**W**	**W**			***L***
from POC2(λ)∘MH∘PH	**W**		**W**			***L***

The loci of L-retraction are shaded. Once again, fusion handles a somewhat complicated prop calc argument with aplomb.

Harmonic completeness can be pursued to a finer level of analysis. Though presented under the heading of onset *inventories* (71), a significantly stronger result is obtained in the actual proof given in OTCIGG:165. The statement of inventory completeness runs like this: if a segment of a certainty sonority level shows up as onset somewhere, then a less sonorous segment will also show up as onset … somewhere. But the proof demonstrates that *wherever* any segment-type is optimally parsed as an onset in any form, there also – in the exact same position — will a less sonorous segment-type inevitably be onset-parsed.

Schematically: if we have an input of the shape /…$[\lambda]_k$…/, with the λ as the k^{th} segment and destined for onset position, then in the exactly corresponding input /…$[\tau]_k$…./, which differs only in that the occurrence of λ mentioned above as $[\lambda]_k$ is replaced by τ, we can be certain that $[\tau]_k$ will also be parsed into an onset.[63] This is stronger because it tells how every segment of a certain type is parsed in every relevant form, rather than making summative remarks about the whole of the emergent inventory.

To approach the argument, we first map out the territory. Let $p = \langle p_1 \rightarrow p_2 \rangle$ be an optimal candidate containing λ as the k^{th} segment in p_1, denoted $[\lambda]_k$, with its output correspondent appearing as an onset in p_2. Since *p* is optimal, it is strictly better than any violationwise-distinct candidate $x = \langle p_1 \rightarrow x' \rangle$. This gives us a collection of ERCs of the form $[p \succ x]$.

Let $q = \langle q_1 \rightarrow q_2 \rangle$ be a candidate which is just like *p*, except that $[\lambda]_k$ is replaced with τ, as is its surface correspondent, if any. For *q* to be optimal, it must be better than any of its alternatives $y = \langle q_1 \rightarrow y' \rangle$. Here we have ERCs of the form $[q \succ y]$.

What we need to establish, in the broadest terms, is that every ERC of $[q \succ y]$ type is entailed by some collection of ERCs of the $[p \succ x]$ type, taken with whatever other ranking conditions Universal Grammar provides. This will ensure that the optimality of *p* entails the optimality of *q*. (If *p* beats all its competitors under some ranking, then *q* must also do the same.)

---

[63] My appreciation for the value of specifying such correspondences in this kind of argument has been increased by discussions with Bruce Tesar. For detailed analysis and relevant proposals, see Tesar 2006 and Tesar, in prep.

The simple structure of the relation between *p* and *q* allows us to match up the ERCs in a way which allows the argument to proceed straightforwardly.

Let's notate the key relation a little more explicitly. For q_1, let's write $\tau(p_1)$, meaning the form arrived at by replacing the k^th^ segment of p_1 with τ. Similarly, we'll write $\tau(p_2)$ for q_2, the form arrived at by replacing the *output correspondent* of $[\lambda]_k$ with τ; and we'll write $\tau(p)$ for whole candidate *q*.

We extend this notation to *x* in the obvious way, so that $y = \tau(x) = \langle\tau(p_1)\to\tau(x')\rangle$, where the last term signifies a form which is like x′ except that the output correspondent of $[\lambda]_k$, if any, has been replaced by τ.[64]

Now we can match up not only the candidates *p* and $q=\tau(p)$, but every candidate in both sets, any candidate *x* based on input /p_1/ with its corresponding $\tau(x)$ based on input /q_1/$=\tau(p)$. We set out to show that $[p\succ x]$ entails $[\tau(p)\succ\tau(x)]$, for every *x*. This will exhaust the collection of suboptimal competitors to $\tau(p)$, and thus establish that $q = \tau(p)$ is optimal when *p* is.

A first, deck-clearing observation: replacing λ with τ has no effect whatsoever on the way many constraints evaluate the forms.The constraint ONS, for example, is insensitive to the λ/τ difference. Thus, *p* and $\tau(p)$ perform equally on ONS, as do *x* and $\tau(x)$, so that ONS assigns the same comparative value to $[p\succ x]$ as to $[\tau(p)\succ \tau(x)]$. From the point of view of entailment, such constraints present no problems.[65] The same holds true for all faithfulness constraints under consideration, since none see λ as distinct from τ. The focus of argument must fall on those constraints that respond differently to *p* and $\tau(p)$, to *x* and $\tau(x)$. These will only be the peak and margin constraints, which are of the markedness subspecies and specifically mention λ and τ.

The argument falls into 3 separate subcases (OTCIGG:165), depending on how *p*'s competitor *x* treats underlying $[\lambda]_k$. Candidate *p*, of course, parses it as an onset. Recall that under our restricted assumptions about Gen, λ cannot change its feature composition.

Case 1. $[\lambda]_k$ is parsed as a **margin** (onset or coda) in candidate *x*.

Case 2. $[\lambda]_k$ is parsed as a **peak** (syllable nucleus) in candidate *x*.

Case 3. $[\lambda]_k$ is **deleted** in candidate *x*.

In each case, we need to determine the effects of τ-for-λ substitution on the relation between the ERCs $[p\succ x]$ and $[\tau(p)\succ \tau(x)]$, with an eye to showing that the second is entailed.

Case 1 is perhaps the easiest. The class of suboptima at issue are those in which the k^th^ segment is parsed as a margin; it is of course always an onset in both optima. The following table lays out the violation effects of τ-for-λ substitution in the k^th^ segment. The alphabetic variables in the cells represent numbers of violations. Mnemonically, we designate the competitor x_{mar}.

(76) **Margin-parsing suboptima** (Case 1)

MAR	*M/λ	*M/τ	*P/τ	*P/λ
p	a	b	c	d
x_{mar}	m	n	f	g
$\tau(p)$	a–1	b+1	c	d
$\tau(x_{mar})$	m–1	n+1	f	g

Note the values of *p* and $\tau(p)$: these remain throughout the entire discussion. Columns will be shaded when they are irrelevant to the argument at hand.

---

[64] We avoid the final natural extension, to $\tau[p\succ x]$ for $[q\succ y]$, merely to maintain visual emphasis in citation of cases.

[65] Because W→W, e→e, L→L.

The violations patterns (shortly to be explicated) reveal that here the τ-transformation has *no effect* on the comparative values. The peak hierarchy, shaded, plays no role at all, because the τ-transformation doesn't change its assessments. The same is true of the margin hierarchy, with a slight arithmetical twist.

A comparative value responds to the *difference* in violations between the two compared candidates. In the case of *M/λ, the τ-transformation decrements both optimum and competitor by one violation, removing one λ-nucleus; for *M/τ, it increments both by one, adding one τ-margin. The optimal/suboptimal comparison remains constant (under translation, as it were). The argument boils down to the observation that

$$a - m = (a-1) - (m-1)$$
$$b - n = (b+1) - (n+1)$$

where the left-hand-side calculation corresponds to the assessment of $[p \succ x_{mar}]$ and the right-hand-side to that of $[\tau(p) \succ \tau(x_{mar})]$.

Conclusion: for this class of suboptimal competitors, the ERCs required to make *p* optimal are *exactly* those required for τ(*p*). If we satisfy them for *p*, we've done τ(*p*) as well.

Let's now justify the claimed violation patterns in more detail. By hypothesis, $[\lambda]_k$ is parsed as a margin in both *p* and *x*. The substitution of τ changes no peaks, hence we ignore the peak hierarchy. Consider first the constraint *M/λ in the margin hierarchy. We lose one λ-margin in the desired optimum τ(*p*) and one in the suboptimal competitor τ(*x*). This lowers their violation scores on *M/λ equally. Whatever relation holds between *p* and x_{mar} on *M/λ — whether they have the same number of violations (*e*), or whether there is more for one than the other (W,L) — that relation must be the same as that holding between τ(*p*) and $\tau(x_{mar})$.

The other relevant margin-sensitive constraint is *M/τ and it behaves the same way. Substituting τ for λ produces one new τ-margin in both the desired optimum $q=\tau(p)$ and one in its competitor $y=\tau(x_{mar})$, raising the *M/τ score in both by one violation. This has no effect on the comparative values assigned to ERC $[\tau(p) \succ \tau(x_{mar})]$, which simply copy those of ERC $[p \succ x]$.

These considerations mean that ERCs $[p \succ x_{mar}]$ and $[\tau(p) \succ \tau(x_{mar})]$ must be *identical*. If the first holds in any grammar, so must the second. Thus, whenever *p* beats all alternatives in which its k^{th} segment is margin-parsed, so must $q = \tau(p)$. Case 1 is closed.[66]

Suppose now that the losing competitor *deletes* the k^{th} segment (case 3). We write x_{del} to name this candidate.

The key observation is this: the τ-transformation has no effect on the evaluation of the suboptimal candidates. It makes no change at all in their *output*s, there being no correspondent of $[\lambda]_k$ in x_{del} to replace. Since all the relevant peak and margin constraints are markedness constraints, evaluating only outputs, the suboptima x_{del} and $\tau(x_{del})$ must fare identically on them.

---

[66] We can arrive at the same conclusion from different perspective, arguing horizontally, as it were, rather than vertically. The optimum *p* is closely matched with its competitor x_{mar} in the treatment of the k^{th} segment, which is parsed as a margin in both. Since the k^{th} segment's featural composition is controlled for, i.e. the same in both candidates, the violation of *M/λ it induces in *p* matches a violation it induces in *x*, and these violations can't have anything to do with determining the value assigned to the comparison by *M/λ. (They 'cancel', in the terminology of the Cancellation Lemma, OTCIGG:164, ex. (216), which validates the inference. Putting it arithmetically, if r is the number of *M/λ violations outside $[\lambda]_k$ in p, and s the number in $x_{mar,}$, then the whole-form totals for each are r+1 and s+1; but (r+1)−(s+1)=r+s.) The k^{th} segment in τ(*p*) has nothing to do with how thecomparison with τ(*x*) fares on *M/λ either. Both $[p \succ x_{mar}]$ and $[\tau(p) \succ \tau(x_{mar})]$ must receive the same value on *M/λ, since the τ-transformation has no effects outside the k^{th} segment. A similar analysis works for *M/τ, arguing back from $\tau(x_{mar})$ to x_{mar}. This line of reasoning is essentially that of OTCIGG:165.

By contrast, the desired optima p and $\tau(p)$ do differ, since the τ-transformation decrements the number of λ-margins and increments the number of τ-margins, just as described in the first case. Checking entailment is a matter of sorting out these effects.

Here's a generic violation table for the competitions involving deletional suboptima.

(77) **Deletional Suboptima – generic violation patterns**

DEL	*M/λ	*M/τ	*P/τ	*P/λ
p	a	b	c	d
x_{del}	m	n	f	g
$\tau(p)$	a–1	b+1	c	d
$\tau(x_{del})$	m	n	f	g

The relations between p and $\tau(p)$ are as above, but x_{del} and $\tau(x_{del})$ are violationwise identical.

We want to show that entailment holds between $[p \succ x_{del}]$ and $[\tau(p) \succ \tau(x_{del})]$. This means that every constraint must adhere to certain relations between comparative values in antecedent and consequent, as in (50). Entailment follows the scale L→ e→W (including self-entailment). There is no point in worrying about cases where L is assigned in the antecedent ERC; whatever value the consequent takes, entailment will stand. But antecedent *e* must take *e* or W in the consequent, and antecedent W must take W. To establish entailment we must scrutinize the cases where the antecedent values are *e* and W, and ascertain how the consequent behaves.[67]

Applying this wisdom to the case at hand, let's concern ourselves first with *M/λ and suppose that $[p \succ x_{del}]$ earns *e* there.This means that m=a in table (77), so that the relevant column comes to look like this.

(78) $[p \succ x_{del}]$ is e on *M/λ

DEL	***M/λ**	*M/τ	*P/τ	*P/λ
p	a	b	c	d
x_{del}	a **e**	n	f	g
$\tau(p)$	a–1	b+1	c	d
$\tau(x_{del})$	a **W**	n	f	g

It is evident that $[\tau(p) \succ \tau(x_{del})]$ earns W here, since the loser's score a is greater than (*i.e.* worse than) the optimum's a–1. Entailment is secure. Now suppose that the antecedent gets W. This means that x_{del} has more violations than p, say a+h for some positive integer h.

(79) $[p \succ x_{del}]$ is W on *M/λ

DEL	***M/λ**	*M/τ	*P/τ	*P/λ
p	a	b	c	d
x_{del}	a+h **W**	n	f	g
$\tau(p)$	a–1	b+1	c	d
$\tau(x_{del})$	a+h **W**	n	f	g

Once again, $[\tau(p) \succ \tau(x_{del})]$ earns W on *M/λ because the desired loser's score is worse, and entailment is secure.

---

[67] In 2-valued logic, T antecedent requires a T consequent. This forms the basis of arguments about entailments in boolean logic. Here the third value (*e*) introduces a second case to consider.

Let's turn our attention to *M/τ. Here the effects are less attractive, because of the way the optima relate. As always, the τ-transformation introduces a new τ margin in $\tau(p)$, a bad thing in the myopic eyes of *M/τ. We pursue the consequences as before by examining the *e* and W cases in the antecedent.

First, the *e*-antecent.

(80) $[p \succ x_{del}]$ is *e* on *M/τ

DEL	*M/λ	***M/τ**	*P/τ	*P/λ
p	a	b	c	d
x_{del}	m	b **e**	f	g
$\tau(p)$	a–1	b+1	c	d
$\tau(x_{del})$	m	b **L**	f	g

Here $[\tau(p) \succ \tau(x_{del})]$ earns L on *M/τ . Entailment fails, because $e \nrightarrow L$. And the problem may repeat if the antecedent carries a W. This means that x_{del} does *worse* than p on *M/τ by having a violation score greater than p's, call it b+h, for some positive constant h.

(81) $[p \succ x_{del}]$ is W on *M/τ

DEL	*M/λ	***M/τ**	*P/τ	*P/λ
p	a	b	c	d
x_{del}	m	b+h **W**	f	g
$\tau(p)$	a–1	b+1	c	d
$\tau(x_{del})$	m	b+h **e,W**	f	g

If it happens that h=1, we're stuck with $[\tau(p) \succ \tau(x_{del})]$ being awarded *e* while $[p \succ x_{del}]$ gets W, killing entailment, because $W \nrightarrow e$.

Waiting in reserve to rectify the situation is the margin hierarchy. When we fuse it in, it becomes clear that combining it with $[p \succ x_{del}]$ yields the desired entailment. We write **v** for whatever comparative value $[p \succ x_{del}]$ happens to obtain from *M/τ.

(82) **MH to the rescue**

	*M/λ	***M/τ**
MH	***W***	***L***
$[p \succ x_{del}]$	**e,W**	**v**
$[p \succ x_{del}] \circ$ MH	**W**	$\mathbf{v} \circ \boldsymbol{L} = \mathbf{L}$
$[\tau(p) \succ \tau(x_{del})]$	**W**	**L,e,W**

As desired, the last row is entailed in any grammar in which p is optimal (second row) and the margin hierarchy is respected (row 1).

In short, the margin hierarchy liberates us from having to worry about *M/τ as long as *M/λ awards W to the τ-transformed comparison. We have verified that this happens. The case of the deletional suboptima is handled by *M/λ.

The final remaining case turns out to crucially involve both hierarchies, but it falls rapidly to the kind of argument pursued here. The relevant class of suboptima is defined by the shared property of parsing the kth segment as a *peak*. The first step is to compute the generic violation system.

(83) **Peak-parsing suboptima** (Case2)

PK	*M/λ	*M/τ	*P/τ	*P/λ
p	a	b	c	d
x_{pk}	m	n	f	g
$\tau(p)$	a–1	b+1	c	d
$\tau(x_{pk})$	m	n	f+1	g–1

The optima behave exactly as above; only the suboptima have changed. The margin hierarchy sees no difference between x_{pk} and $\tau(x_{pk})$, since the τ-transformation affects only their peak count. Substituting τ for λ in the kth position in x_{pk} increments *P/τ by one violation, and decrements *P/λ similarly. To calculate the effects, let us consider only those cases where $[p \succ x_{pk}]$ earns *e* or W: the competitor x_{pk}'s violation values can be written as the optimum's value plus a constant greater than or equal to zero. (We compress the 0 and positive cases here.)

(84) **Peak-parsing suboptima** — All things considered. $h_i \geq 0$.

PK	*M/λ	*M/τ	*P/τ	*P/λ
p	a	b	c	d
x_{pk}	$a+h_1$ **e,W**	$b+h_2$	$c+h_3$ **e,W**	$d+h_4$
$\tau(p)$	a–1	b+1	c	d
$\tau(x_{pk})$	$a+h_1$ **W**	$b+h_2$	$c+h_3+1$ **W**	$d+h_4-1$

We have shaded the lower ends of each hierarchy, because we know that they are irrelevant to the entailment situation whenever the upper ends assess W of the τ-transformed suboptima.[68] This is easily seen to be the case,[69] and we're done.

After detailed calculations like these, it is useful to register motivating ideas and broad-stroke understanding against the actual outcome. The leading assumption is surely the hypothesis, or view, that lower sonority ought to mean greater affinity for the onset, the canonical consonantal position. The Margin hierarchy specifies a generalized version of this idea, leaving the onset/coda distinction to other constraints. In the present case, we start by assuming that we have a form in which a relatively-higher sonority element is optimally parsed as an onset. We then deduce that anything less sonorous, in the exact same position, would also be onset-parsed. The Margin hierarchy is necessary to achieve this result, as we'd expect. Strikingly, though, there are cases where it is irrelevant and cases where it is not sufficient.

When the suboptimal competitor parses the target segment as a margin, as in (76), the mere fact that *something* is onset-parsed in that position in the optimum ensures that *anything* will be better parsed as an onset there — better than any other margin-parsers. We found that the ERCs associated with λ and with τ are identical in this case: whatever certifies the superiority of the desired optimum in one case will do just as well for the other. (We don't know what it is — in a particular candidate, it could even be a constraint from the Margin hierarchy reacting to

---

[68] Fusion with the relevant peak and margin hierarchy ERCS drives them to L, which entails anything.

[69] Because for any a,c, it's true that $a-1 < a+h_1$ and $c < c+h_3+1$.

substructures somewhere else in the form.) But the implicational relation between the two cases, which runs in both directions, has nothing to do with the marginal sonority. This is because the optimum and its competitor are matched, in the crucial position, for both margin-hood and sonority. It can't be a difference in the quality of this particular position that sways the choice, because there are no relevant differences. Uniformly substituting a new segment into that position in both optimum and competitor maintains the sonority match; it may slide the total scores on a given margin hierarchy constraint up or down, but always by the same amount in both candidates. This result is very much what we'd expect: *all other things being equal*, onset parsing should be universally preferred to any other kind of margin-parsing, i.e. as coda.

For deletional suboptima, the Margin hierarchy is required to maintain the entailment relation, as shown in (80)-(82). It is precisely, and only, the less sonorous segment's superiority as a margin that carries the day, ensuring entailment. This is because the suboptimal competitors are identical in the λ-containing base case and its τ-substituted image. Since the suboptima are held constant, the extra value of τ as a margin simply adds an advantage that guarantees that successful onset-parsing of the λ-form cannot fail to force similar treatment of the τ-form.

The most interesting case, perhaps, is provided by the suboptima in which the targeted segment is parsed as a peak, a syllable nucleus. Here the Margin hierarchy is necessary, but isn't in itself sufficient to ensure entailment. Without the Peak hierarchy the argument would not go through. This conclusion runs contrary to expectations nurtured by approaches that generalize over inventories and also intuitions implanted by noninteractive theories in which a particular virtue is sufficient in itself to license well-formedness. But the result is natural here. Anti-structuralist to its core, the OT of OTCIGG is not directly about inventories, but is concerned entirely with derivation, with finding the optimal match between a particular input and a particular output. Consequently, the focus falls on what can happen in suboptimal forms to the very τ that is optimally onset-parsed. One of the inevitable options is that it be a nucleus; and, to obtain the desired entailment it must be decided whether the nuclear parse amounts to a comparative flaw or a comparative advantage. Were it —heaven forbid —worse to parse λ as a peak than τ, entailment would fail. In this hypothetical case, the constraint $*P/\lambda$ would be allowed a say in the entailment reckoning. Since *decreasing* the number of λ-peaks is regarded favorably by the constraint, the comparative value tips from *e* to L, and from W to *e* in certain circumstances. We repeat the configuration from the last column of (84), where the comparative situation went uncalculated because the Peak hierarchy renders it irrelevant. But if the Peak hierarchy were upended or abolished, it would be irrelevant no more.[70]

(85) **Behavior of $*P/\lambda$**, when $h_4 \geq 0$.

PK	$*P/\lambda$
p	d
x_{pk}	$d+h_4$ ***e*, W**
$\tau(p)$	d
$\tau(x_{pk})$	$d+h_4-1$ **L, *e*,W**

Entailment dies for two reasons. In a ranked hierarchy, $*P/\lambda$ could easily be in a position where it does no harm to $[p\succ x_{pk}]$, granting it an *e*, but itself kills off $[\tau(p)\succ\tau(x_{pk})]$ via L (when $h_4 = 0$,

---

[70] Calculations: if $[p\succ x_{pk}]$ gets e, then $h_4 = 0$. This gives $d-1$ for the violation score of $\tau(x_{pk})$, yielding L. If $h_4 = 1$, then $[p\succ x_{del}]$ earns W while $[\tau(p)\succ\tau(x_{pk})]$ earns *e*. Recall that entailment is never endangered when the antecedent has L, so we only consider the *e*- and W-antecedent cases.

invoking $e \nrightarrow L$). And even when *P/λ potentially decides [$p \succ x_{pk}$] by assigning it W, it could disastrously fail to decide [$\tau(p) \succ \tau(x_{pk})$] (with h_4 =1, invoking $W \nrightarrow e$), failing to protect it from any antagonistic constraint lurking lower in the hierarchy. In the real world these dangers are averted by the higher-ranking *P/τ, which awards W to [$\tau(p) \succ \tau(x_{pk})$] in the cases at hand and thereby shields it from any judgments offered by *P/λ.

The situation is summarized in the following purely comparative representation:

(86) **Potential Entailment Disaster**. NB: $e \nrightarrow L$, $W \nrightarrow e$, $\alpha \nRightarrow \beta$

PK		*Mλ	*M/τ	*P/τ	*P/λ
α	$p \succ x_{pk}$	…	…	*e*,W	***e*, W**
β	$\tau(p) \succ \tau(x_{pk})$	…	…	W	**L, *e*,W**

Even when $[\lambda]_k$ is onsetted in *p*, the structurally-parallel candidate τ(*p*) could lose to a counter-analysis $\tau(x_{pk})$ in which $[\tau]_k$ is nuclear — but for the intervention of the peak hierarchy.

It is thus an irremovable fact of the theory's logic that the harmonic completeness of onsets must reflect in some way the role played by margin and peak hierarchies. We reckon it an advantage of the approach through the ERC calculus that the logical dependencies are unavoidably classified and laid bare. Such transparency enables the analyst and improver to actually analyze and improve, rather than to merely envision and persuade through cases, guess-work, attitude.

10. Evolution of the ERC

Here error is all in the not done,
all in the diffidence that faltered . . .
—Canto LXXXI

Where did the ERC come from and where is it going? The arguments developed in OTCIGG, which lie behind those just discussed, exploit a logic that is often parallel to that used here, as we have emphasized. The explicit basis is different, though, and both the similarities and the departures might be missed in a hasty flip-through.

The observation essential to the OTCIGG approach is that a flaw shared by competitors cannot distinguish them. Comparing two candidates on any single constraint, we may discard any violation marks they have in common: *canceling* marks one for one, until the candidate doing better on that constraint has none left. Applied to every constraint, this would yield a mark-cancelled version of the original violation tableau, in which, for each of the two candidates involved, the constraints (still) carrying violations are exactly those disfavoring it in the comparison; those constraint that favor it in a comparison and those that don't care are shown as violation-free. Here's an example that gives the cases, with uncancelled marks shown as bullets.

(87) **Uncancelled violation tableau**

	C_1	C_2	C3
p	**	**	**
x	****	**	*

(88) **Mark-cancelled version**

	C_1	C_2	C_3
p			●
x	●●		

If *p* is the desired winner in the competition, then constraints that *must be dominated* are any identified as assigning it uncancelled marks (here, only C_3). Those constraints that *may dominate* — i.e. those from which the necessary dominator may be drawn — assign uncancelled marks to the desired loser *x* (here, only C_1).

The validity of this calculation is established by the Cancellation/Domination Lemma. (The candidate names in the citation have been changed to reflect this paper's conventions.)

(89) **Cancellation/Domination Lemma** (CDL)

> Suppose two candidates *p* and *x* do not incur identical sets of marks. Then $p \succ x$ iff every mark incurred by *p* which is not cancelled by a mark of *x* is dominated by an uncancelled mark of *x*.
>
> OTCIGG: p. 261, ex, (238), repeating p.154, ex. (192)

In present terms, an 'uncancelled mark' of *x* is a constraint awarding W to $[p \succ x]$, and an uncancelled 'mark incurred by *p*' is a constraint awarding L to the comparision. The Cancellation/Domination Lemma is transmuted into the Elementary Ranking Condition: 'every L is dominated by some W', which is equivalent to 'some W dominates every L' under linear ordering. The mark-cancelled representation just given would come out like this:

(90) **Comparative tableau**

	C_1	C_2	C_3
$[p \succ x]$	**W**		**L**

In OTCIGG, mark cancellation uses each optimum-suboptimum pair under consideration to induce its own partition of the constraint set into the possibly dominating, the necessarily dominated, and the irrelevant.[71] After classifying the constraints with respect to one winner-loser pair, its job is done, and all further work is conducted in terms of the constraints thus identified.

There can be no such thing as a general mark-cancelled tableau; the process only works for pairs of candidates. It crucially includes the transformed optimum in the results, and the optimum may be differently mark-cancelled in the context of different competitors. On any given constraint, one loser's victory can be paralleled by another loser's defeat. The following example, simplified from Prince 2002b:4-5, illustrates this phenomenon.

---

[71] Compare the notions W-set and L-set of Prince 2002a:1. The W-set of an ERC is the set of constraints contributing W's to it; similarly for the L-set. (These correspond to the 'winner marks' and 'loser marks' of mark cancellation theory.) Together they classify the constraints that must be worried about in a ranking argument. They form a 'polarity' in the sense of semantics of relevance logic (Dunn 1986:189).

(91) **Uncancelled and cancelled versions of two comparisons**

Raw	C_1
q	*
x	
y	**

Cancelled	C_1
q	●
x	

Cancelled	C_1
q	
y	●

The desired optimum *q* cannot be shown as both bulleted and unbulleted in the same cell.

The move to ERC theory transforms mark cancellation from process to representation. Individual violation profiles, which record only shortcomings (observable but not interpretable in isolation), serve as the basis for generating a formal object which marks both advantages and disadvantages (as well as neutrality), and thereby incorporates and displays the central notion of comparison that the theory is based on. The comparative tableau — and specifically the row or 'vector' of comparative values — is recognized as an entity unto itself, one with its own laws of logical relation (W-extension, L-retraction) and combination (fusion, negation, and their relatives). With this in hand, it is no significant distance to the indispensible notion of ERC entailment, essential for investigating implication and impossibility, equally essential for understanding of the structure of ranking requirements (Prince 2002a, Brasoveanu & Prince 2005). We have a tool designed for the uses we need to put it to.

Two questions about comparative representation arise in the context of linguistics as it is practiced. The first is inevitable, given the rhetorical history of various intergroup struggles. Is it a *mere notational variant*? As emphasized throughout, comparative representation is qualitatively different in character from violation lists. A computation relates them, which loses all information about the number of violation marks and obtains information about the specific better-than/worse-than/same-as relationships they imply.

The second question is a subtler version of the first: is it worth the intellectual overhead? We already have violation lists and violation tableaux with constraints in domination order and the '!' annotation to mark extinction of candidates. Do we really need more equipment? The reality, of course, is that in setting out to do OT at all, we already face the requirement for more. If you wish to determine the ranking requirements imposed by data, you must make the calculation that determines the distribution of W's and L's. If you wish to find and dismiss the harmonically bounded, you must compare candidates. If you wish to understand the patterns of implication and impossibility that follow from a posited constraint system, you must deal with the logic of constraint domination and connect it with the 'better than' relation in data. The choice of not-doing-more is simply not on offer. The choice lies in how you are going to do it.

11. Possible Grammars and Possible Grammatical Systems

A *formal grammar* is any linear ordering of the constraint set. This definition constructs the notion purely on the constraint side. Without reference to candidates, forms, or relations between forms, it guarantees us that a grammar will have the key property of resolving all conflicts so as to produce one optimal violation profile from every candidate set. But we cannot tell, looking at the constraints alone, which of the formal grammars define the same language.

Over on the candidate side, we have the *grammatical system*: a comprehensive choice of optima — a 'language', understood to be a collection of input-output relations, chosen from every input. Looking at any such collection of candidates, drawn from each candidate set, we cannot tell without reference to the constraints whether it is a *possible* grammatical system, a human language, or just a random gathering of forms.When the choice yields a system that the theory generates, the typical result is several, possibly many, extensionally-equivalent formal grammars.[72] The empirical force of the everpresent plurality of formal grammars lies in the implications and impossibilities that underwrite the many-to-one mapping between linear orders and grammatical systems. Much of the content and predictiveness of linguistic theory lies right there.

As opposed to a grammatical system broadly construed — an output for every input — a *possible* grammatical system is one sanctioned by a ranking. Any ranking that suffices to produce a given grammatical system will conform to a certain set of requirements, specifically those imposed by the ERCs derived from *all* the optimum-suboptimum comparisons in the system. Despite the impressive number of such comparisons (as unlimited as the number of candidate sets and candidates), the entire collection of distinct ERCs on *n* constraints is most assuredly finite, and so any subset must be. A *possible grammatical system* may therefore be finitely characterized in terms of ERCs and ERC entailment.

This speaks to a natural question that arises from the use of ERCs in arguing *about* grammatical systems. Are they merely instrumental, tools to be put away when the work is done? The distinction between a *formal grammar* and a *grammatical system* suggests the contrary. The ultimate goal and stopping point of analysis is not, and cannot be, to find one formal grammar for the data, a hierarchy that works. The patterns of explanation — as we have seen — lie in the necessary and sufficient ranking conditions, and in their relations across candidate sets. Typically, many of the relations in any given hierarchy will be artifacts of linearization, and scrutinizing a linear hierarchy, including its patterns of filtration, cannot reveal its structure. The primary aim of analysis must therefore be to arrive at the relevant ranking conditions — a set of ERCs that guarantees the desired optima.

It is worth noting, in this context, that work on learning supports the primacy of the ERC set. The first learning idea that springs to mind is that, given data, the learner should guess a grammar and forget the data; given more data, the learner ought to modify the current grammar accordingly. From each encounter with data, on this view, the learner retains only a grammar hypothesis. But this has proved to be far too brittle, artifact-ridden, and information-poor a conception to support the learner's progress. Tesar 1997ab determines that what the learner needs to keep is actually (in our terms) the set of ERCs generated by the data, from which a hierarchized grammar may be produced easily when needed for filtration purposes; this finding informs subsequent work in the area. More recently, the learner's ERC collection has been termed the 'support'; there is no reason to assume that the support is ever jettisoned. This means that the learner continues with, and ends up with, an ERC set. Prince & Tesar 2002 examine in some detail the apparently ineradicable difficulties in trying to replace the support with a formal grammar (linear or stratified) in the course of learning. In this line of research, then, what the learner keeps must be *knowledge*, or as close as can be gotten to it, and the intrusion of artifacts indistinguishable from truth is unwelcome.

There is also a purely formal sense in which the ERC set is irreplaceable. Diagramming partial orders is an accessible, entirely rigorous way to present them, and has played a role in OT work since at least McCarthy & Prince 1993/2001:60. But the set of linear constraint orders consistent with a grammatical system is *not* in general equivalent to a partial order. It follows that such 'Hasse' diagrams cannot be trusted to encode the rankings that define a grammatical system.

---

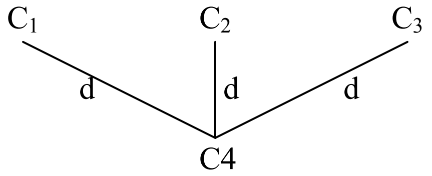
[72] As noted at the outset, this is virtually inevitable on broad formal grounds, given the thoroughgoing requirements of symmetry that both constraints and candidates must meet if (anything like) each of the *n*! grammars is to correspond to a distinct grammatical system.

This may be seen in the simplest of examples: consider a system defined by one ERC of the form (W,W,L).The temptation is to construe this as a partial order in which both C_1 and C_2 are ordered above C_3. But (W,W,L) also admits linearizations $C_1 \gg \mathbf{C_3} \gg \mathbf{C_2}$ and $C_2 \gg \mathbf{C_3} \gg \mathbf{C_1}$ in which the first and second constraints (bearing W) are separated by the third (the lone L). The only partial order that would support both of these is one in which all constraints are unordered with respect to each other. The natural way to extend the Hasse diagram is to differentiate among the arcs by some form of labeling. It won't work to mark all 'disjunctive' arcs with one code, however, because disjunctivity is local to a specific ERC. Consider the two ERC system

α: (W,W,e,L)
β: (e, W,W,L)

We might imagine that we could mark the arcs with '*d*', say, indicating that C_1 and C_2, as well as C_2 and C_3 are disjunctive with respect to C_4.

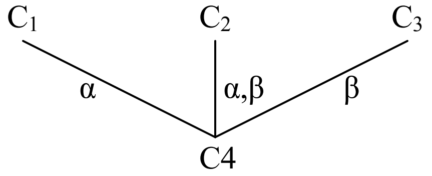
(92) **Inadequate representation of system {α,β}**



This diagrams fails to distinguish the system from (W,W,W,L). This allows $\mathbf{C_1} \gg \mathbf{C_4} \gg C_2 \gg C_3$, which fails to satisfy ERC β = (e,W,W,L).

We must therefore distinguish between the arcs coming from α and those coming from β, labeling with the ERC name.

(93) **Representing {α,β}**



The rule of linearization is now that at least one arc bearing a given name must be realized as precedence in a licit order. The order $C_1 \gg C_4 \gg C_2 \gg C_3$ fails to represent {α,β} because neither of the β requirements is respected.

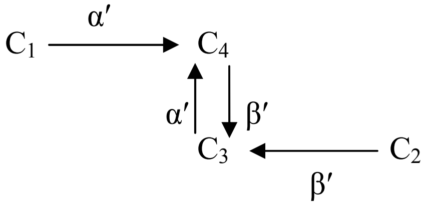
Even this modification falls well short of general adequacy. This may been in a system like the following, which superficially resembles the one just examined:

α′ : (W,e,W,L)
β′ : (e.W,L,W)

This system contains what we might call a 'pseudo-cycle' involving the third and fourth constraints. From α′ we see that C_3 is a potential dominator of C_4. From β′ we that C_4 is a potential dominator of C_3. Since all potential domination relations must be represented in the graph, we must have a directed arc going from C_3 to C_4 and another running back from C_4 to C_3.

The result would look like this:

(94) **Domination relations as a directed, arc-labeled, cyclic graph**



To define the classes of admitted rankings, we must specify that each constraint-node appears once and only once in any linearization, while each label must be satisfied at least once.[73]

We are far from the Hasse Diagram, in which all flow is from top to bottom and there is no need for explicitly directed arcs. The level of complication that can easily be reached (barely hinted at in our example) is such as to eliminate legibility and ease of manipulation. Instead of replacing or obliterating its source, this method of diagramming takes us right back to the ERC set as the canonical representation of ranking requirements.

How, then, can we usefully characterize the key notions *formal grammar* and *possible grammatical system* in ERC terms?

The linear order in a formal grammar has the interesting property that it decides all comparisons between *every* pair of candidates. Though we use it to select optima, it imposes order throughout the entire set of candidates, pervading the suboptimal regions.[74] To see this, observe with Samek-Lodovici & Prince 1999:53 that if a grammar G is taken to be a function producing an optimum from a set of candidates (regardless of what other properties it may have), we can use it to deliver the 'better than' relation. For any pair of distinct violation profiles x,y, put them to the grammar as the entirety of a candidate set. If $G(x,y)=x$, so that x is 'optimal' in the mini-candidate set at hand, then we can say that $x \succ y$, in (provably) the same sense as we have used '$\succ$' above.

Now let G be a linear order on a constraint set. Since x and y are distinct on some constraint, we can legitimately speak of the highest-ranked constraint (existence guaranteed by linearity) that distinguishes them (they do differ). This constraint will decide between them. Indeed, x and y need not even be linguistically-realizable candidates — any two distinct points in the space of violation profiles defined by the constraint set will be ordered by a linear hierarchy in this way.[75]

Moving to ERC-talk, the notion of 'deciding between two candidates' translates exactly as dealing with the comparison between them, either $\alpha = [x \succ y]$ or $\neg\alpha = [y \succ x]$. An ERC set decides

---

[73] Diagrams of similar character are produced by OTSoft, the very useful calculation program developed by Bruce Hayes and his collaborators, available at Hayes's website, http://www.linguistics.ucla.edu/people/hayes/otsoft/. A different approach is reflected in the 'domination graphs'' of Raymond & Hogan 1993.

[74] OTCIGG:82-91, Samek-Lodovici & Prince 1999:53-56. Coetzee 2004 proposes that this order among suboptima, usually assumed invisible, emerges in variation. His formal claim that OT as defined in OTCIGG only imposes an optimum vs. suboptimum distinction is, however, inaccurate.

[75] The profile space of an n-constraint set is a subset, typically proper, of $\mathbb{N}^n = \mathbb{N}\times\mathbb{N}\ldots\times\mathbb{N}$, where $\mathbb{N}$ is the set of nonnegative integers. This exact structure of the space will depend on the constraint set at hand. If C_1, for example, is boolean on candidates and only distinguishes between satisfaction and violation, the dimension of the profile space that corresponds to it can only have two elements: 0,1.

the comparison between distinct *x* and *y* in favor of *x* iff it can be augmented consistently with the ERC [x≻y], but becomes inconsistent when [y≻x] is added to it. This is, of course, precisely equivalent to saying that an ERC set decides the competition between *x* and *y* in favor of *x* iff [x≻y] is *entailed* by the set.

In ERC territory, the linearly-ordered hierarchy G will correspond to a set of ERCs, or more precisely, various logically equivalent ERC sets; pick any one and call it Γ. The finding is that any ERC set Γ that requires a linear order will have the *entailment* property of 'completeness': namely, that for every possible ERC α , Γ entails either α or ¬α.

(95) **Completeness**. An ERC set Γ over a set of *n* constraints Σ is *complete* with respect to the set of all ERCs U on the profile space of Σ iff for every α∈U, Γ entails α or Γ entails ¬α.

Carlos Fasola raises the question of whether the implication involving linearity also runs in the opposite direction: if an ERC set is complete, must it yield a unique linear order? This would be true if we could guarantee that every pair of constraints was rankable: i.e. supported a (W,L) ERC. This is not necessarily the case. A pair of constraints may define their violations so that harmonic bounding orders the candidates, yielding what Prince (1997-2001) calls a 'stringency hierarchy' (on candidates). Consider the following example:

(96) **Stringency Hierarchy on whole Candidates** .

	C_1	C_2
x		
y		*
z	*	*

Here the relation between any two candidates is the same regardless of constraint ranking. Either order will do, and any set of ERCs is complete with respect to it — including the empty set! This is because the empty set entails valid ERCs, and the ERCs created from the pairs here are either valid or invalid, as we show in this semicomprehensive table (the other half being generable by negation).

(97) **ERCs from the Stringency Hierarchy** .

		C1	C2
α	*x≻y*	**e**	**W**
β	*x≻z*	**W**	**W**
γ	*y≻z*	**W**	**e**

The ERC set (W,L) entails α = (e,W). But so does (L,W). No linear order is required here for completeness with respect to {α,β,γ}.

The notion of a *complete ERC set* is as close as we can come in ERC logic to a linear hierarchy, using candidate-based ERCs.

A *grammatical system* differs from a *formal grammar* in that it only decides optima. There is no guarantee that its decision-making capacities extend into the suboptimal regions, and lack of completeness will be common. It's easy to see why this is so. Suppose that in an idealized case we have a system in which two constraints much each dominate a third, i.e. $\{C_1,C_2\}\gg C_3$. This places no restrictions on the ranking of C_1 and C_2, yet is perfectly compatible with a profile space containing suboptimal elements that are distinguished by that relation. Consider this schematic example, in which *x*, *y*, and *z* are all possible optima, and *x* is chosen as the optimum in a particular grammar.

(98) **Among the suboptima**

	C1	C2	C3
x			⊛
y	* **W**	**e**	**L**
z	**e**	* **W**	**L**

The relation between candidates *y* and *z* is *not* decided by this ERC set. Evaluating [*y*≻*z*] yields (L,W,e), which is not entailed by the optimality of *x*, i.e. by (W,e,L) or (e,W,L), the members of {[*x*≻*y*], [*x*≻*z*]}.

We have seen a concrete example with precisely this character, in the pre-V candidate set of the *t,d*-deletion cases, repeated here:

(99) **Pre-V deletion**

IN	OUT	*CPLX	MWd	ONS	PHR	MAXC	*Remark*
1. cost us	a. cost.us	* **W**		*		**L**	faith1
	b. cos.tus		* **W**	**L**		**L**	faith2
	c. cos.us			⊛		⊛	***del***

No relation is established in the deletional grammatical system between the retentive suboptima ⟨a⟩ *cos**t**.us* and ⟨b⟩ *cos.**t**us*. Their relative status follows from the ranking of *CPLX and MWd, which is unconstrained when ⟨c⟩ is optimal. Even when the rest of the deletional system is included, there is still no means of deciding between them.

The notion of a *possible grammatical system* combines candidate-side and constraint-side information, and cannot be delimited without referring to both. In the *t,d*-deletion case, for example, we might ask whether the ERC set of (99) defines a grammatical system. To answer, we must find out if there are candidate sets for which it does not determine an optimum. The pre-pause input provides such a case:

(100) **Comparatively Annotated Tableau: Deletion everywhere**

IN	OUT	*CPLX	MWd	ONS	PHR	MAXC	*Remark*
3. cost##	a. cost.	* **W**			**L**	**L**	faith
	b. cos.				*	*	***del***

Comparing this ERC with those in (99) shows it to be unentailed, therefore independent of them. The following tableau brings them all together. (Note the e↛L relations in the PHR column.)

(101) **Deletion Grammar**

ERC	*CPLX	MWd	ONS	PHR	MAXC
[1ca]	**W**				**L**
[1cb]		**W**	**L**		**L**
[**3ba**]	**W**			**L**	**L**

To define a possible grammatical system, an ERC set must guarantee choice of an optimum from every candidate set. This means that it must entail every ERC that can be created from optimum-suboptimum comparisons. To show that this holds will require grasp of the structure of the candidate sets produced by Gen. In the t,d-deletion case, it is essential to recognize that the pre-pausal environment is analyzed differently by the constraint set than the others, and in particular to notice that its optimum is guaranteed by an unentailed ERC.

The ERC characterization of a grammatical system has proven itself basic to the exploration and understanding of implication and impossibility within a constaint set (§3,§4,§9). The key recognition is that cases on the candidate side that differ in the linguistic structures they involve may nonetheless receive logically-related analyses when connected by ERCs to their constraint-side ranking requirements. The relation may be as straightforward as identity, as we saw above in comparing pre-V and pre-C deletion, which share ERC [1ca], or it may involve subtler entailments of various degrees of complexity.

We may discern here a further use of the ERC calculus. Just as the structure of ERC entailment within a candidate set identifies a finite collection of suboptimal candidates whose defeat ensures optimality — victory over every competitor —, so does ERC entailment within a grammatical system identify a finite collection of *candidate sets* that produces all the ranking requirements needed to define an optimum in every candidate set. If the task of finding such a definitive collection is not brought to completion, a proposed grammar can fail even with respect to readily available data, overlooking items and patterns that bear crucial information. Failure to complete may also be more abstract, leading to nondecision somewhere in the full range of candidate-sets defined by Gen, which may include types not represented in near-to-hand data.

It is natural to approach the task of completion by enumerating the classes of candidate sets distinguishable by the constraint set, and then mining each class for ERCs and testing for sufficiency, aiming to show that all possible ERCs are entailed by the ones thus derived. This will require nontrivial reference to constraints, constraint interaction effects, and the linguistic structure of candidates. The ERC, as a bridge between the constraint-side and the candidate-side, taken with its intrinsic logic, provides the tools for conclusively exploring such relations.

■ ■ ■

12. Appendix: the Aligned-Retention Grammar

Here we provide a view of the grammar producing aligned-retention in all cases.

(102) **Annotated Violation Table for the Constraint Set**

IN	OUT	*CPLX	MWd	ONS	MWd	MAXC	*Remark*
1. cost us	**a. cost.us**	⊛		⊛			**AR**
	b. cos.tus	**L**	* **W**	**L**			nonAR
	c. cos.us	**L**		*		* **W**	del
2. cost me	**a. cost.me**	⊛					**AR**
	b. cos.tme	*	* **W**				nonAR
	c. cos.me	**L**				* **W**	del
3. cost##	**a. cost.**	⊛					**AR**
	b. cos.	**L**			* **W**	* **W**	del

This is the purely comparative format:

(103) **Comparative Tableau**

ERC#	IN	Winner	Loser	*CPLX	MWd	ONS	MWd	MAXC	*Remark*
1ab	cost us	**cost.us**	cos.tus	**L**	**W**	**L**			*nonAR
1ac			cos.us	**L**				**W**	*del
2ab	cost me	**cost.me**	cos.tme		**W**				*nonAR
2ac			cos.me	**L**				**W**	*del
3ab	cost##	cost.	cos.	**L**			**W**	**W**	*del

Remark: [2ab] is useless, so the set of ERCs reduces to this:

(104) **Comparative Tableau**

ERC#	IN	Winner	Loser	*CPLX	MWd	ONS	MWd	MAXC	*Remark*
1ab	cost us	**cost.us**	cos.tus	**L**	**W**	**L**			*nonAR
1ac			cos.us	**L**				**W**	*del
2ac	cost me	**cost.me**	cos.me	**L**				**W**	*del
3ab	cost##	cost.	cos.	**L**			**W**	**W**	*del

From this, the following implicational relations are clear:

(105)

Implication	**Justification**	**As in text**
AR/Pre-V ⇒ AR/Pre-C	[1ac]=[2ac]	¬[1ca] = ¬[2ca]
AR/Pre-C ⇒ AR/Pre-##	[2ac]→[3ab]	¬[3ba] → ¬[2ca]

* * *

References

ROA = http://roa.rutgers.edu

Anderson, A. and N. Belnap. 1975. *Entailment*: *the Logic of Relevance and Necessity*. Vol 1. Princeton University Press. Princeton.

Anttila, A. 2006. Variation in Optimality Theory -1,2,3,4, Bib. Handouts, http://www.stanford.edu/~anttila/teaching/indiana/variation.html

Anttila, A. and C. Andrus. 2006. T-Orders. ROA-873.

Bakovic, E. 2006. Elsewhere effects in Optimality Theory. ROA-844.2.

Beckman, J. 1997. *Positional Faithfulness*. 1998 UMass dissertation. ROA-234.

Brasoveanu, A. and A. Prince. 2005. Ranking and Necessity. Part I. ROA-794.

Clements, G. N. 1991. Place of articulation in vowels and consonants: a unified theory. [1993, as Lieu d'articulation des consonnes et des voyelles: une théorie unifiée. In Bernard Laks and Annie Rialland, eds., collections Sciences du Langage, *Architecture des représentations phonologiques*, pp. 101-145. Paris, France: Editions du CNRS.]

Clements, G.N. and S.J. Keyser. 1983. *CV Phonology*: *a Generative Theory of the Syllable*. MIT Press. Cambridge.

Coetzee, A. 2004. *What it Means to be a Loser: Non-optimal Candidates in Optimality Theory*. Ph.D. Dissertation. ROA-687.

Dunn, J. M. Relevance logic and entailment. In D. Gabbay and F. Guenther, eds. *Handbook of Philosophical Logic. Volume III: Alternatives in Classical Logic*. D. Reidel, Dordrecht.

Green, Tom. 1993. The Conspiracy of Completeness. ROA-8.

Kiparsky, P. 1993. An OT Perspective on Phonological Variation. Handout, http://www.stanford.edu/~kiparsky/Papers/nwave94.pdf.

Lacy, Paul de. 2004. Markedness conflation in Optimality Theory. *Phonology* 21.2:145-199.

Lacy, Paul de. 2006. *Markedness: Reduction and Preservation in Phonology*. CUP.

McCarthy, J. & A. Prince. 1993a. Prosodic Morphology I. As: *Prosodic Morphology: Constraint Interaction and Satisfaction*: ROA-482.

McCarthy, J. and A. Prince. 1993b. Generalized Alignment. In Geert Booij and Jaap van Marle, eds. *Yearbook of Morphology 1993*, pp. 79-153. ROA-7.

McCarthy, J. & A. Prince. 1995. Faithfulness and Reduplicative Identity. ROA-60.

Meyer, R. K. 1975. Chapters 29.3 and 29.12 in Anderson & Belnap 1975.

Prince, A. 1997-2001. Handouts on stringency. http://ling.rutgers.edu/people/faculty/prince.html

- o Stringency and Anti-Paninian Hierarchies. Handout II. LSA Linguistic Institute at Cornell, 23 June- 2 July, 1997
- o Paninian Relations. University of Massachusetts at Amherst, October 31, 1997.
- o [Son of] Paninian Relations. University of Marburg, March 17, 1998.
- o The Special and The General. Penn Linguistics Colloquium, February 26, 2000.
- o Invariance under Re-Ranking. WCCFL XX, held at USC, February 24, 2001.

Prince, A. 2002a. *Entailed Ranking Arguments*. ROA-500.

Prince, A. 2002.b. Arguing Optimality. ROA-562.

Prince, A. & P. Smolensky 2004 (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell: Malden-Oxford-Carlton. Also available as ROA-537.

Prince, Alan, & B. Tesar. 2004. Learning phonotactic distributions. In Rene Kager, Joe Pater, and Wim Zonneveld, eds., *Fixing Priorities: Constraints in Phonological Acquisition*. Cambridge University Press. 245-291. Also as ROA-353 and RUCCS-TR-54.

Raymond, W. and A. Hogan. 1993. The Optimality Interpreter (C program), and A User's Guide to the Optimality Interpreter. ROA-130.

Samek-Lodovici, V. 2001. Crosslinguistic Typologies in Optimality Theory. In G. Legendre, J. Grimshaw, and S. Vikner, eds. 2001, *Optimality-Theoretic Syntax*. MITPress.

Samek-Lodovici, V. and A. Prince. 1999. Optima. ROA-363.

Samek-Lodovici, V. and A. Prince. 2005. Fundamental Properties of Harmonic Bounding. ROA-785.

Smolensky, P. 2006. Harmonic completeness, local conjunction, and feature domains. Ch. 14 in Smolensky & Legendre 2006.

Smolensky, P. and G. Legendre. 2006. *The Harmonic Mind: from neural computation to Optimality Theoretic grammar.* Vol 2: *Linguistic and Philosophical Implications*. MIT Press: Cambridge.

Tesar, B. 1997a. Multi-recursive constraint demotion. ROA-197.

Tesar, B. 2006. Faithful Contrastive Features in Learning. *Cognitive Science* 30:5.

Tesar, B. In prep. Transparent maps.

Tesar, Bruce. 1997b. Using the mutual inconsistency of structural descriptions to overcome ambiguity in language learning. In *Proceedings of the North East Linguistic Society* 28, Pius N. Tamanji and Kiyomi Kusumoto, eds., p. 469-483. Amherst, MA: GLSA, University of Massachusetts.

Yip, Moira (1983) "Some Problems of Syllable Structure in Axininca Campa," in Peter Sells and Charles Jones, eds., *Proceedings of NELS* 13, Graduate Linguistic Student Association, University of Massachusetts, Amherst, pp. 243-251.

■ ■ ■