Doing OT in a Straitjacket

Jason Eisner (Penn / Rochester) — UCLA, 14 June 1999
jeisner@linc.cis.upenn.edu

1. A question that could drive you crazy

What constraints does OT allow? i.e., What is the substance of the theory?

(1) Some clearly *bad constraints* (but what makes them bad?):
  a. Palindromic: The candidate reads the same backwards as forwards.
  b. FtQuint: Feet are quintary (5 syllables or moras).
  c. MemberOf(*a, aardvark, aardvarks, aardwolf, aardwolves, Aaron* ...): Candidate must be in the specified set of surface forms.
  d. MatchesOutputOfSPE: The output matches the result of applying Chomsky & Halle (1968) to the input.

(2) Some clearly *okay constraints* (but what makes them okay?):
  a. Clash-ATR: Low vowels may not bear the atr feature.
  b. Onset: Every syllable must start with a consonant.

(3) Some questionable constraints, by the standards of derivational phonology:
  a. FtBin: Feet are binary (2 syllables or moras).
  b. Align-L(Foot, PrWd): The sum of all distances from left edges of feet to the left edge of the PrWd is minimized. (For consequences see (34).)
  c. Half the constraints that first-year phonology students make up.

Reasons to try to formalize OT, rather than allowing *ad hoc* English constraints:

(4) a. Results in an explicit, falsifiable theory of UG
  b. Simplifies that theory, exposing formal similarities among constraints
  c. Enables computational work (e.g., Eisner 1997b)
    (tools for linguists; algorithms for generation, parsing, acquisition; theorems on expressive power)
  d. Constrains linguistic description
  e. Aids descriptive work by providing well-motivated and well-formalized constraints and representations
    (many constraints given informally in the literature, including GA, do not specify how to count violations in all circumstances)

The formalization sketched in this talk is called **OTP**—OT with primitive constraints.

(5) Identifying such core constraints is at the center of the OT program:
  "The danger, therefore, lies in ... clinging to a conception of Universal Grammar as little more than a loose organizing framework for grammars. A much stronger stance, in close accord with the thrust of recent work, is available ... *Universal Grammar can supply the very substance from which grammars are built: a set of highly general constraints, which, through ranking, interact* to produce the elaborate particularity of individual languages." (Prince & Smolensky 1993, p. 198)
  (see also Smolensky 1995, Green 1994)

2. Traveling the Web in search of truth

Suppose we had a set **Con** of core constraints for phonology—simple mechanisms that could be used to build up all the basic phonological phenomena. What would it look like?

Ask: What formal devices are regularly used by constraints in the literature?

(6) a. NasVoi (Itô, Mester, & Padgett 1996)
    "Every nasal segment must be linked to some voicing feature."
  b. Onset (Prince & Smolensky 1993) } (equivalent)
    Align(σ, L, C, L) (McCarthy & Prince 1993)
    "Every syllable must begin with (be left-aligned with) some consonant."
  c. *Common thread:* "Every ... some."
    $\forall\alpha, \exists\beta$ such that α and β stand in such-and-such local relationship.

If we allow α and β to be edges (as one option), we only need one kind of local relationship—*temporal coocurrence*:

(7) The primitive **implication** family.
  $\boxed{\alpha \rightarrow \beta}$ means: $\forall\alpha, \exists\beta$ such that α and β coincide temporally.

(8) Rewrite (6):
  a. $nas \rightarrow voi$: $\forall nas, \exists voi$ such that nas and voi coincide temporally.
  b. ${}_\sigma[\rightarrow {}_C[$: $\forall {}_\sigma[, \exists {}_C[$ such that ${}_\sigma[$ and ${}_C[$ coincide temporally.

Thus we can regard alignment as "edge licensing." (Or licensing is "feature alignment.") We can also mix references to edges and interiors:

(9) $F \rightarrow]_\mu$: Every foot must cross a mora boundary. (No degenerate feet.)
  (= Min-2m: Green & Kenstowicz 1995)

Like GA, primitive implication is formal rather than substantive:
  Onset: ${}_\sigma[\rightarrow {}_C[$, NoOnset: ${}_\sigma[\rightarrow {}_V[$, Coda: $]_\sigma \rightarrow]_C$, and NoCoda: $]_\sigma \rightarrow]_V$ are all equally easy to express using this family. So as in other theories, UG must still state that Onset and NoCoda are strongly preferred by human grammars (just as it must state any universal rankings). (The dispreferred constraints may still be useful: e.g., Hammond 1995 proposes a NoOnset constraint for stressless syllables. See Green 1994 on metaconstraints.)

McCarthy & Prince (1993) have previously noted that alignment plays a unifying role, and have suggested that it's *the* core mechanism for all of phonology:

(10) a. "These examples only hint at the generality of the phenomenon to be explored

here, which extends to include all the various ways that constituents may be enjoined to share an edge in prosody and morphology. Data like these have been given widely disparate treatments in the literature ..." (p. 1)

b. "Taken together with $\overline{\text{X}}$-like restrictions on immediate domination and interpreted within the appropriate theory of constraint satisfaction, *GA provides a mechanism for completely specifying a class of formal languages that*, when substantive parameters are set, *ought to be all-but-coextensive with possible human languages.*" (p. 2)

A second constraint family:

Above, we unified feature licensing and alignment.

The opposite of feature licensing is feature clash.

The opposite of alignment is disalignment, i.e., edge clash.

(11) a. *[low, ATR] (Cole & Kisseberth 1994)

"Low features are incompatible with ATR features."

b. NONFINALITY = *ALIGN($PrWd$, R, F, R) (e.g., Buckley 1995)

"Prosodic words may not be right-aligned with feet."

(12) The primitive **clash** family.

$\boxed{\alpha \perp \beta}$ means: $\forall\alpha$, $\nexists\beta$ such that α and β coincide temporally. [cf. (7)]

Equivalently: $\forall\alpha\ \forall\beta$, α and β are temporally disjoint.

(13) Rewrite (11):

a. *low* $\perp$ ATR: All *low* and ATR features are temporally disjoint.

b. $]_{PrWd} \perp]_F$: Each $]_{PrWd}$ does not coincide with (fall on) any $]_F$.

Again, this formulation suggests we can mix edges and interiors, and we can:

(14) $F \perp {}_M[$: A foot may not cross a morpheme boundary.

(= TAUTO-F, Crowhurst 1994)

(In fact, (14) is more plausible than Crowhurst's formulation, $*\,{}_F[\sigma\ {}_M[\ \sigma\ ]_F$. It would be surprising to find a language that crucially blocked ${}_M[$ only where Crowhurst states, while still allowing it to interrupt a syllable or a ternary foot.)

Null hypothesis: These two families of local $\boxed{\text{primitive constraints}}$—implication and clash—are the *only* ones needed.

$\boxed{\begin{array}{l}\alpha \to \beta \text{ says that } \alpha\text{'s attract } \beta\text{'s.}\\ \alpha \perp \beta \text{ says that } \alpha\text{'s repel } \beta\text{'s.}\end{array}}$

3. What representations are being constrained?

The primitive constraints constantly refer to edges, and never refer to association lines. So they are easiest to interpret if we assume that *ŋk* is represented as in (15b),

not (15a). This representation is inspired by Optimal Domains Theory (Cole & Kisseberth 1994) and Correspondence Theory (McCarthy & Prince 1995).

(15) a.

```
  voi
nas/
 |/
 C   C
  \ /
  vel
```

(Goldsmith 1990)

b.

$_{voi}[$	$]_{voi}$		
$_{nas}[$	$]_{nas}$		
$_C[$	$_C][_C$		$]_C$
$_{vel}[$			$]_{vel}$

⟶⟶timeline⟶⟶

(The timeline is really just an ordered set of edge brackets. Thus, only horizontal order matters in the drawing above. Ignore spacing and vertical order.)

c. Easy to see that (15b) violates the progressive voicing constraint $]_{voi} \perp {}_C[$.

(16) Key characteristics of the new representation:

a. Constituents float along a *timeline*.

Example constituents: *nas* (autosegmental), μ (prosodic), `x` (stress mark), *Stem* (morphological), *H*-domain (feature domain)

b. The timeline is continuous, not divided into segments.

c. *All* constituents have width and edges. Thus we can refer naturally to the edges of syllables (or morphemes) whose segmental features are scattered across multiple tiers and perhaps shared with other syllables (cf. Itô & Mester 1994).

d. For autosegments with width, such as [*nas*], think of phonetic gestures. (15b), which begins with simultaneous ${}_{nas}[$ (= lower the velum) and ${}_{voi}[$ (= begin vibration of the vocal folds). The primitive constraints can only affect the *order* of bracket edges; it is up to the phonetic component to determine actual durations.

e. Association or Correspondence of two constituents is indicated by having them overlap. (Independently proposed by Bird & Klein 1990.) E.g., the velar gesture in candidate (15b) spans both consonants.

f. No need for faithfulness constraints on the insertion, deletion, or relocation of association lines (cf. Kirchner 1993, Myers 1994, Féry 1994).

g. No need for (inviolable) well-formedness constraints against gapping or crossing of associations (cf. Kirchner 1993, Féry 1994, Oostendorp 1995).

h. No need for Correspondence indices.

(17) Where do these representations come from? The behavior of **Gen**:

a. *Hypothesis:* **Gen** can't do anything fancy, like palindromes—nothing that the primitive constraints couldn't also handle. So for convenience, let's make Gen as simple as possible, and let undominated constraints clean up the huge unbridled candidate set that results.

b. Gen places constituents *freely* along the continuous timeline.

That is, as far as Gen is concerned, brackets may land anywhere.

Conditions such as the prosodic hierarchy are enforced by undominated primitive constraints, not by Gen.

c. However, Gen requires that edge brackets come in matched pairs.

d. Gen also does not allow distinct constituents of the same type (e.g., two syllables or two *lab* autosegments) to overlap. (Elements on the same tier never link to each other.)

e. Gen is free only with regard to output material. It is forced to place a copy of the input material into every candidate, on its own tier, for purposes of I-O Correspondence. (Cf. Containment (Prince & Smolensky 1993), Strict Consistency Constraint (Polgardi 1995).)

(18) Because the timeline is continuous rather than divided into segments, brackets can fall in mid-segment:

a. Contour tones:

$_H[\qquad\quad]_H$
$\qquad\quad_L[\qquad\quad]_L$
$\qquad_V[\quad\ ]_V$

```
    H
  . \   L
 .   \ / .
     vowel
```

b. Geminates (long vowels are similar):

$_\sigma[\qquad\qquad\qquad\qquad]_\sigma\qquad\quad]_\sigma$
$_C[\quad\ ]_C\quad _C[\qquad\ ]_C$
$\qquad_V[\quad\ ]_V\qquad _V[\quad\ ]_V$

```
σ   σ
/|\ /|
C V C V
```

4. Donning the straitjacket

(19) Formal statement of the primitive constraint families:

a. $\alpha \rightarrow \beta$: Each α temporally overlaps some β.
Scoring: Each α without a β incurs one violation mark.

b. $\alpha \perp \beta$: Each α temporally overlaps no β.
Scoring: Each overlap incurs one violation mark.

(20) What can α and β be?

a. *Edges* such as $_{low}[$ or $]_{low}$.

b. *Interiors* such as *low*.
Denote only the interior of a constituent, *without its edges.*
Thus, *low* and ATR do not overlap here: $_{ATR}[\qquad]_{ATR}$ / $_{low}[\qquad]_{low}$
I.e., the above candidate satisfies *low* $\perp$ ATR but violates *low* $\rightarrow$ ATR.

c. *Conjunctions* and *disjunctions* as in (21).
(Dispreferred in analyses, on grounds of their greater complexity—they refer to more features.)

(21) Occasionally, must allow the following generalized forms of (19). I propose to limit conjunction/disjunction to these configurations only.

a. (α_1 and α_2 and ...) $\rightarrow$ (β_1 or β_2 or ...)
Scoring: Violated once by each set of objects $\{A_1, A_2, \ldots\}$ of types $\alpha_1, \alpha_2, \ldots$ respectively that all overlap on the timeline and whose intersection does not overlap any object of type $\beta_1, \beta_2, \ldots$.

b. (α_1 and α_2 and ...) $\perp$ (β_1 and β_2 and ...)

Scoring: Violated once by each set of objects $\{A_1, A_2, \ldots, B_1, B_2, \ldots\}$ of types $\alpha_1, \alpha_2, \ldots, \beta_1, \beta_2, \ldots$ respectively that all overlap on the timeline.
(Could also be notated: $\alpha_1 \perp \alpha_2 \perp \cdots \perp \beta_1 \perp \beta_2 \perp \cdots$.)

Each violation mark is still triggered individually by a bad *local* condition in the candidate, e.g., a moment on the timeline when certain edges are present and others are not.

Note that some constraints require crisp alignment of edges ($_x[\rightarrow {}_y[$), while others are weaker and require only overlap ($x \rightarrow y$), allowing spreading. (Cf. the violable CRISPEDGE constraint of Itô & Mester (1994).)

5. Input and output

I-O Correspondence (between input and output features): Signaled by alignment between input and output tiers.

Correspondence relations with and without spreading:

(22) a. $_{voi}[\quad]_{voi}$ / $_{\underline{voi}}[\quad]_{\underline{voi}}$ — Perfect faithfulness

b. $_{\underline{voi}}[\quad]_{\underline{voi}}$ — Violates MAX-IO (PARSE): $\underline{voi} \rightarrow voi$

c. $_{voi}[\quad]_{voi}$ — Violates DEP-IO (FILL): $voi \rightarrow \underline{voi}$

d. $_{voi}[\qquad\qquad]_{voi}$ / $_{\underline{voi}}[\quad]_{\underline{voi}}$ — Like (a), this spread version satisfies PARSE & FILL, which only require overlap. Spreading may be required to satisfy some other constraint. On the other hand, various constraints can be invoked against spreading: either $voi \perp]_{\underline{voi}}$ or $]_{voi} \rightarrow]_{\underline{voi}}$ or $]_{\underline{voi}} \rightarrow]_{voi}$ could be used to block (d). (Cf. Yip, 1994:21,fn. 11, on MSEG vs. *Insert Structure)

Thus, the timeline mechanism unifies Correspondence relations with autosegmental associations. Both are encoded by overlap on the constituent timeline. This fleshes out a proposal of McCarthy & Prince (1995):

(23) "The re-casting of autosegmental association in terms of correspondence relations may be expected to have consequences for the analysis of tonal, harmonic, and related phenomena. We do not explore these ideas here, though they are clearly worth developing." (p. 22)

Traditional view: The input to Gen is an underlying form drawn from the lexicon. Gen(x) produces all candidates whose input tiers *exactly encode* x.

Broader view (Tranel 1994): The input to Gen is an abstract morpheme (or sequence thereof). Gen(x) produces all candidates whose input tiers are *compatible with* x. So Gen picks the underlying form(s).

Competing candidates may then differ to some extent even on their input tiers:

(24) a. *phonologically conditioned allomorphy:* candidates try different allomorphs on the input timeline, and the constraints decide what works best. (Tranel's "free suppletion")

b. *floating tones and features:* the lexicon specifies only that H falls somewhere on the input. Different candidates try different locations for it in the input. The output may or may not correspond.

c. *floating morphemes, templatic morphology:* morphology specifies the order of underlying segments within each morpheme, but lets the morphemes overlap so that their segments intermix freely on the input tiers. These segments may or may not be preserved in the output.

d. *epenthesis* ($\underline{CC} \Rightarrow CVC$): The lexicon doesn't specify whether input segments are adjacent, so Gen can create candidates that push them apart:

$$\begin{array}{llll} & {}_V[& &]_V \\ {}_C[&]_C & {}_C[&]_C \\ {}_{\underline{C}}[&]_{\underline{C}} & {}_{\underline{C}}[&]_{\underline{C}} \end{array}$$

syncope ($\underline{CVC} \Rightarrow CC$) is represented similarly:

$$\begin{array}{llll} {}_C[&]_C & {}_C[&]_C \\ {}_{\underline{C}}[&]_{\underline{C}} & {}_{\underline{C}}[&]_{\underline{C}} \\ & {}_{\underline{V}}[& &]_{\underline{V}} \end{array}$$

But for good phonological reasons, we'd like other constraints to regard as adjacent the two surface C's in the latter case (and the two $\underline{C}$'s in the former). See §12 for a very general solution.

6. This straitjacket fits pretty comfortably

This section illustrates how all the types of primitive constraints are ubiquitous across different areas of phonology.

My apologies in advance for any errors or mischaracterizations in these lists. Some of these translations to OTP are not exact, but appear to act correctly on the data in the papers cited. Also, note that sometimes there is more than one way to paraphrase a constraint.

("ROA" citations (http://ruccs.rutgers.edu/roa.html) not further listed in the bibliography.)

Key to unfamiliar notation:

feat	version of feature on output tier
$\underline{feat}$	version of feature on input tier (underline denotes "underlyin' " material)
μ_s	strong mora, containing onset and nucleus (Zec 1988).
μ_w	weak mora, containing coda if any (Zec 1988). (One could also use explicit constituents Ons, Nuc, Coda.)
x	a 2ndary stress mark over a stress-bearing unit (first layer of the grid)
X	a word-primary stress mark (second layer of the grid)
Seg	segmental root node (alternatively, C or V), as distinguished from morphological root *Root*

Some implication constraints from the literature.

(25) **"Same edge" implication:**

a. **Features**

1. $]_{raised} \rightarrow]_{upper}$ — Align[R][U]. Bradshaw ROA-93j.

b. **Prosody**

1. $]_{PrWd} \rightarrow]_\sigma$ — Align: Wd] = σ]. Myers, ROA-6.
2. $]_F \rightarrow]_{\mu_w}$ — Iambic Quantity: In a rhythmic unit (W S), S is heavy. Hung, ROA-24.
3. $]_{PrWd} \rightarrow]_{\mu_w}$ — Align-H: Align(*PrWd*, R, heavy syllable, R). Kager, ROA-70.
4. $\mathtt{x}[\rightarrow {}_F[$ — Foot-Form (trochaic): If there is a head, it is on the L. Hung, ROA-9. Trochaic: Align($\acute{\sigma}$, L, Foot, L). Kager, ROA-35.
5. ${}_F[\rightarrow \mathtt{x}[$ — Align(Ft, L; Head(Ft), L). Bermudez-Otero, ROA-136.
6. $]_{PrWd} \rightarrow]_{\mathtt{X}}$ — Final-Str: Align(domain, R, $\acute{\sigma}$, R). Kager, ROA-35.
7. $]_F \rightarrow]_\sigma$ — Fill: Respect the usual prosodic hierarchy, without catalexis. Inkelas, ROA-39. (Take catalexis to be ${}_F[\ {}_\sigma[\cdots]_\sigma \cdots]_F$, and assume another constraint $]_F \perp \sigma$.)

c. **Feature-prosody interaction**

1. ${}_F[\rightarrow {}_C[$ — Align(Ft, L, Onset): The left edge of a foot must always be aligned to the onset of the first syllable in the foot. Goedemans, ROA-26. (Assume we also have ${}_F[\rightarrow {}_\sigma[$.)
2. ${}_C[\rightarrow \mathtt{x}[$ — NoOnset: Stressless syllables do not have onsets. Hammond, ROA-58.
3. ${}_H[\rightarrow {}_{PrWd}[$ — Align(H tone, L, PWd, L). Myers, ROA-6.
4. $]_{\mu_s} \rightarrow]_{son}$, et al. — HNuc: A higher sonority nucleus is more harmonic than one of lower sonority. Féry, ROA-34, following P&S 1993.
5. $]_{\mu_s} \rightarrow]_V$ — Project($\overline{\overline{\mathrm{N}}}$, V): Nucleus must be a vowel. Oostendorp, ROA-84.
6. ${}_\sigma[\rightarrow {}_{A_0}[$ — Strong Onset: Syllables begin with a closure A_0. Bakovic, ROA-96.

7. $(\,]_\sigma$ and $]_{hi}\,) \rightarrow\,]_{back}$ — *. . . i$]_\sigma$. Kenstowicz, ROA-103.
8. $(\,]_{low}$ and $]_\sigma\,) \rightarrow\,]_{\mathtt{x}}$ — No [a]: [a] is not allowed in unstressed open syllables. Kager, ROA-93a.
9. $(\,]_{hi}$ and $]_\sigma\,) \rightarrow (\,]_{\mathtt{x}}$ or $]_{back}\,)$ — No [i]: [i] is not allowed in unstressed open syllables. Kager, ROA-93a.

d. **I-O relationships**

1. ${}_{Hdom}[\ \rightarrow\ {}_{\underline{H}}[$ — Left-Hd: The leftmost tone bearer of a tone span must be a head. Myers, ROA-6.
2. $]_{\underline{ATR}} \rightarrow\,]_{ATRdom}$ — BA-rt: Align(Anchor-s, R; [atr]-domain, R). Cole & Kisseberth, ROA-22.

e. **Morphophonology**

1. $]_{Plural} \rightarrow\,]_{son}$ — Son]Pl: Plurals end in a sonorant. Golston & Wiese, ROA-100.
2. ${}_{\underline{M}}[\ \rightarrow\ {}_F[$ — Morpheme-Foot-Left: Align(Morpheme, L, Foot, L), where "a *single* violation is assessed for every morpheme which does not meet this requirement." Crowhurst, ROA-19. See also Kager, ROA-35; Bermudez-Otero, ROA-136.
3. ${}_{\underline{Root}}[\ \rightarrow\ {}_{PrWd}[$ — Align-WD: Align(root, Left; PrWd, Left). Cohn & McCarthy, ROA-25.
4. ${}_{\underline{Root}}[\ \rightarrow\ {}_\sigma[$, etc. — Align(Root, σ; L,R): "Align root morpheme boundaries with syllable bondaries at both edges." Yip, ROA-14.
5. ${}_{Red}[\ \rightarrow\ {}_F[\ ,\ ]_{Red} \rightarrow\,]_F$ — Red = Foot. ROA-16. Carleton & Myers, ROA-16. (Also need $Red \perp {}_F[$.)

(26) **"Opposite edge" implication:**

a. **Features**

1. $]_{lax} \rightarrow {}_{\mu_w}[$ — Project(lax, $\overline{\mathrm{N}}$): Lax vowels are followed by additional weight (coda consonant or 2nd half of a diphthong). Oostendorp, ROA-84.
2. ${}_{\mu_w}[\ \rightarrow\,]_{lax}$ — Project($\overline{\mathrm{N}}$, lax): Only lax vowels are followed by additional weight (as if tense vowels bore their own). Oostendorp, ROA-84.
3. $(\,]_{vel}$ and ${}_C[\,) \rightarrow (\,]_{cont}$ or $]_{voi}\,)$ — No kC. Bradshaw, ROA-93j.

b. **Prosody**

1. $]_{\mathtt{x}} \rightarrow {}_\mu[$ — Rhythm: A stressed element must be followed by an unstressed element. Hung, ROA-9. (Also need $]_{\mathtt{x}} \perp {}_{\mathtt{x}}[$.)
2. $(\,]_\sigma$ and ${}_\sigma[\,) \rightarrow (\,]_{\mathtt{x}}$ or ${}_{\mathtt{x}}[\,)$ — NoLapse: No adjacent unstressed syllables. Anttila, ROA-63.

3. $(\,]_\sigma$ and ${}_\sigma[\,) \rightarrow (\,]_{\mathtt{x}}$ or ${}_{\mathtt{x}}[$ or $]_F$ or ${}_F[\,)$ — Lapse: Adjacent unstressed syllables are separated by a foot boundary. Green, ROA-45.

c. **I-O relationships**

1. ${}_H[\ \rightarrow\,]_{\underline{H}}$ — Local: An output TBU bearing tone t must be adjacent to [input] TBU b, where b [also] bears t. Bickmore (credited to Myers), ROA-161. (Only right spreading actually appears. Note the variation ${}_H[\ \rightarrow (\ {}_{\underline{H}}[$ or $]_{\underline{H}}\ )$.)

d. **Morphophonology**

1. ${}_{\underline{Affix}}[\ \rightarrow\,]_{PrWd}$ — Align-SFX: Align(Affix, L, *PrWd*, R). McCarthy & Prince, ROA-7.

(27) **"Interior" implication:**

a. **Features**

1. $rd \rightarrow back$ — *Round* $\rightarrow$ *Back*. Cole & Kisseberth, ROA-98.
2. $nas \rightarrow voi$ — NasVoi. Itô, Mester, & Padgett, ROA-38; Yip, ROA-81.
3. $V \rightarrow ATRdom$ — WSA-lf: Align([atr]-dom, L; Word, L). Cole & Kisseberth, ROA-22. (This gets the correct, gradient effect of spreading as far as possible.)
4. $nas \rightarrow Seg$, etc. — Features like *nas* surface only if linked to a (faithful or epenthetic) segmental root. Zoll, ROA-143.
5. $ATR \rightarrow ATRdom$ — Not explicitly mentioned in Cole & Kisseberth, ROA-22, but clearly needed there.
6. $\sigma \rightarrow (\,H$ or $L\,)$ — Max-ET: Every TBU must have a correspondent tone. McCarthy & Prince (1995). Spec(Tone): Every TBU has a tone. Zoll, ROA-143, after Prince & Smolensky (1993).
7. $V \rightarrow (\,front$ or $round$ or $low\,)$ — Color: A vowel is [front] or [round] if it is [-low]. Kirchner, ROA-4.
8. $C \rightarrow (\,cor$ or lab or $dors\,)$ — $C \rightarrow F_C$: A [+cons] root dominates a consonantal place feature. Oostendorp, ROA-84.
9. $(\,ATRdom$ and $V\,) \rightarrow ATR$ — Express: Express[ATR]. Cole & Kisseberth, ROA-22.

b. **Prosody**

1. $\mu \rightarrow \sigma$ — Parse μ: Every mora must be parsed into a syllable. Myers, ROA-6.
2. $\mu_w \rightarrow \mathtt{x}$ — Weight-to-Stress: Heavy syllables are stressed. Hung, ROA-9 (following Prince 1990).
3. $Seg \rightarrow \sigma$ — Parse(Root): Every root node must be associated with a syllable or mora.

c. **Feature-prosody interaction**

1. $\sigma \rightarrow H$ — Fill(σ): A syllable must be associated with a [high tone. Myers, ROA-6.
2. $V \rightarrow Nuc$ — $V \rightarrow \sigma$: A vowel must be a syllable head. Green, ROA-8.
3. $Nuc \rightarrow son$ — $\sigma \rightarrow R$: A syllable head must be at least a resonant. Green, ROA-8.
4. $round \rightarrow$ (*back* or *stress*) — MaV(Pro) (Marked Vowel (Prominent)): Umlauted vowels fall in prominent syllables. Féry, ROA-34.
5. x $\rightarrow$ (*lo* or *hi* or *front* or *back*) — Non-Head(ə): Stressed schwa is prohibited. Cohn & McCarthy, ROA-25.

d. **I-O relationships**

1. $\underline{H} \rightarrow H$, etc. — Parse(T): A tone must be parsed. Myers, ROA-6.
2. $\underline{lab} \rightarrow lab$, etc. — MaxPl: Parse underlying place features. Lombardi, ROA-105. Max, McCarthy & Prince 1995.
3. $lab \rightarrow \underline{lab}$, etc. — *Ins(F): Do not insert features. Kirchner, ROA-4. Dep, McCarthy & Prince 1995.
4. $\underline{\mu} \rightarrow \mu$ — WeightIdent: If an input vowel is bimoraic, then so is the correspondent output vowel. Pater, ROA-107. See also WeightIdent, Alderete, ROA-131.
5. $\underline{\text{x}} \rightarrow$ x — StressIdent: Parse lexical stress. Pater, ROA-107. Head-Max: Alderete, ROA-131 (from McCarthy 1995).
6. ($\underline{\text{x}}$ and *Affix*) $\rightarrow$ x — Head-Max$_{\text{Affix}}$: Specializes Head-Max to affixes. Alderete, ROA-131.
7. (*Seg* and x) $\rightarrow$ $\underline{Seg}$ — Head-Dep: Every segment contained in a prosodic head in S_2 [output] has a correspondent in S_1 [input]. Roberts-Kohno, ROA-93k.
8. ($\underline{nas}$ and x) $\rightarrow$ *nas*, etc. — HeadSyll-Max(F): No features are deleted from (parsed?) segments in the head syllable. Yip, ROA-159.
9. ($\underline{\mu}$ and x) $\rightarrow \mu$, etc. — Head-Wt-Ident: No lengthening or shortening of stressed syllables. Alderete, ROA-131.
10. $H \rightarrow$ ($\underline{H}$ or $\underline{L}$) — TPFaith: Preserve tonal prominence profile. Tranel, ROA-72; Zoll, ROA-143.

e. **Morphophonology**

1. $\underline{MWd} \rightarrow \bar{\text{X}}$ — HeadProj: $_{MWd}[$ ...Head(PWd) ... $]_{MWd}$. A lexical head must project a prosodic head: every MWd constituent must include a stressed vowel. (A strengthened replacement for Lx≈Pr.) Kennedy, ROA-139.
2. $\underline{M} \rightarrow PrWd$ — MorPa: At least one element of a morpheme is incorporated into a prosodic word. Oostendorp, ROA-84.
3. $\underline{Root} \rightarrow F$ — Ft-Root: The root must overlap with a foot. Buckley, ROA-93c.

(28) **"Mixed" implication:**

a. **Features**

1. $upper \rightarrow {}_{\mu}[$ — Minimal Tone Association (MTA): [+upper] must be linked to more than one TBU. Bradshaw, ROA-93j.
2. ($]_{A_0}$ and $_{A_f}[$) $\rightarrow pal$ — NoAff: Disallows non-palatal affricates. Bakovic, ROA-96.
3. ($]_C$ and $_C[$) $\rightarrow$ (*cor* or *dors* ...) — Contact: Coda should share place with the following Onset [if any]. Kenstowicz, ROA-30.
4. ($]_{nas}$ and $_C[$) $\rightarrow voi$ — *NÇ: No nasal – voiceless obstruent sequences. Pater, ROA-160.
5. (*voi* and $_C[$) $\rightarrow]_{nas}$ $\gg$ ... — ... $\gg$ No-NC-Link, Itô, Mester, & Padgett, ROA-38.

b. **Prosody**

1. $F \rightarrow {}_{\mu}[$ — Min-2m: A metrical foot contains at least two moras. Green & Kenstowicz, ROA-101.
2. $PrWd \rightarrow {}_{\sigma}[$ — Disyll: The left and right edges of the PrWd, must coincide, respectively, with the left and right edges of *different* syllables. Kager, ROA-70. (Also need $_{PrWd}[\rightarrow {}_{Seg}[$, $]_{PrWd} \rightarrow]_{Seg}$.)
3. ($]_{\sigma}$ and $_{\sigma}[$) $\rightarrow$ ($]_F$ or $_F[$ or F) — Parse-2: One of two adjacent stress units should be parsed by a foot. Kager, ROA-35. Parse-Adj-Syll. Alderete, ROA-94.

c. **Feature-prosody interaction**

1. ($_F[$ and $\underline{Root}$) $\rightarrow {}_C[$ — FtOnset$^{\{rt\}}$: Align(Ft that is in root, L, C or Root, L). Buckley, ROA-56.
2. (V and $_{\mu_w}[$) $\rightarrow low$ — Lower: Long vowels are low. $V_{\mu\mu} \rightarrow [Low]$. Cole & Kisseberth, ROA-98.

d. **I-O relationship**

1. $(\,H$ and ${}_{\sigma}[\ )\ \rightarrow\ ]_{\underline{H}}$ T-Bin: A tone span can have at most one non-head (in a domain); limits spread to one syllable from underlying tone. Myers, ROA-6.

Some clash constraints from the literature.

(29) **"Same edge" clash:**

a. **Prosody**

1. $]_{\mathtt{x}} \perp\ ]_{PrWd}$ — *Final Stress. Anttila, ROA-63. Non-Fin($\acute{\sigma}$). Cohn & McCarthy, ROA-25. Cf. Rhythm, Hung, ROA-9.
2. $]_{F} \perp\ ]_{PrWd}$. — NonFinality: Feet should not be word-final. Ní Chiosáin, ROA-89 (credited to Spaelti as WeakEdge(P-Cat)), et al.

b. **Feature-prosody interaction**

1. ${}_{\sigma}[\ \perp\ {}_{nas}[$ — *Ons/N. Smolensky, ROA-86 (following Prince & Smolensky 1993).
2. $]_{lax} \perp\ ]_{\sigma}$ — Project(lax, $\overline{\text{N}}$): Lax vowels are followed by additional weight (coda consonant or 2nd half of a diphthong). Oostendorp, ROA-84.
3. $]_{obs} \perp\ ]_{\mu_w}$ — *ObsNuc. Pater, ROA-107.
4. $(\ ]_{C}$ and $]_{\sigma}\ )\ \perp\ ]_{lab}$ — CodaCond: Syllable-final consonant may not have place features. Lombardi, ROA-105.

c. **I-O relations**

1. ${}_{H}[\ \perp\ {}_{\underline{H}}[$ — *Align(H,L)-I/O: High tone in output must not left-align with its position in input. Bickmore, ROA-161.
2. $(\ ]_{PrWd}$ and $]_{\mu_w}\ )\ \perp\ ]_{V}$ — Free-V: *PrWd*-final vowels must not be parsed. So final heavy syllables are CVC, not CVV. Kager, ROA-70.

d. **Morphophonology**

1. $]_{\underline{M}} \perp\ ]_{low}$ — *a]: No low vowel in a morpheme-final open syllable. Kager, ROA-93c.
2. ${}_{H}[\ \perp\ {}_{\underline{M}}[$ — *Align(H, L, Source Morpheme, L) with no violation by distance. Bickmore, ROA-161.

(30) **"Opposite edge" clash:**

a. **Features**

1. $]_{H} \perp\ {}_{H}[$ — OCP: *FF, where F is a parsed [output] feature specification. "Furthermore, we will consider two tones to be adjacent if they are associated by parsed associations with adjacent tone bearers" (so domains are unnecessary). Myers, ROA-6.
2. $]_{son} \perp\ {}_{voi}[$ — *rg: No sonorant-voiced clusters. Ní Chiosáin, ROA-89.
3. $(\ ]_{nas}$ and ${}_{C}[\ )\ \perp\ ]_{voi}$ — *NÇ: No nasal – voiceless obstruent sequences. Pater, ROA-160.

4. $(\ ]_{vel}$ and $]_{cont}\ )\ \perp\ {}_{lab}[$ — No VelCont Lab: No sequence of a velar continuant before a labial. Bradshaw, ROA-93j.
5. $(\ ]_{nas}$ and ${}_{C}[\ )\ \perp\ voi$ — No-NC-Link. Itô, Mester, & Padgett, ROA-38.

b. **Prosody**

1. $]_{\mathtt{x}} \perp\ {}_{\mathtt{x}}[$ — *Clash: No adjacent strong beats on the grid. Kager, ROA-35. NoClash. Anttila, ROA-63. Cf. Rhythm, Hung, ROA-9.
2. $]_{F} \perp\ {}_{F}[$ — *FtFt: Feet must not be adjacent. Kager, ROA-35.

(31) **"Interior" clash:**

a. **Features**

1. $voi \perp gl$ — *[voiced, gl]: No implosives. Buckley, ROA-57.
2. $tense \perp low$ — *TENSE-low: No tense low vowels. Benua, ROA-74.
3. $phar \perp dor$ — *Mid (no mid vowels): *[Phar, Dor]. Alderete, ROA-94.
4. $hi \perp low$ — Non-occurrence of +hi and +low. Kirchner, ROA-4.
5. $Seg \perp Word$ — *Structure(Root). Myers, ROA-6.
6. $H \perp Word$ — *Struct(A): There must be no association. Myers, ROA-6.
7. $low \perp Word$ — *[low]. Oostendorp, ROA-84 (following Prince & Smolensky 1993).

b. **Prosody**

1. $\sigma \perp PrWd$ — Monosyllabicity: The fewer syllables, the better. Noske, ROA-109. *Struc(σ): No syllables. Zoll, ROA-143.

c. **Feature-prosody interaction**

1. $\mu_w \perp (\ gl$ and $\ldots\ )$ — Coda-h: A /h/ may only occur in an onset. Oostendorp, ROA-84.

(32) **"Mixed" clash:**

a. **Features**

1. $hi \perp {}_{Seg}[,\ lo \perp {}_{Seg}[$ — *Mult-Height: No multiply linked height features. Kirchner, ROA-4.
2. $front \perp {}_{\underline{front}}[$, etc. — *Spread: Do not insert association lines.
3. $RdDom \perp {}_{HiDom}[$, etc. — Uniformity: The (round-)harmony domain must be monotonic: high or low. Cole & Kisseberth, ROA-98. (Cf. parasitic harmony.)
4. $(\ ]_{V}$ and ${}_{V}[\ )\ \perp hi$, etc. — NoLongVowel: Two adjacent vocalic roots may not be linked to the same material (but diphthongs are allowed). Oostendorp, ROA-84.

b. **Prosody**

1. $F \perp {}_M[$ — TAUTOMORPHEMIC-FOOT: $*_F[\sigma\ {}_M[\sigma]_F$. Crowhurst, ROA-19.
2. $\mu_s \perp {}_{Seg}[$ — *BRANCH(S)μ. Walker, ROA-142.
3. $F \perp {}_\sigma[$, etc. — UNARITY: A prosodic category *p* contains no more than one of the next lower prosodic category p-1. A. Green, ROA-115.
4. ${}_F[\perp \sigma$ — SYLLINT: Syllable integrity (violable). Everett, ROA-163.
5. $\sigma \perp (\ ]_C$ and ${}_C[\ )$ — *COMPLEX: Only one element can be in onset or coda position.

c. **Feature-prosody interaction**

1. $C \perp]_\sigma$ — GEMINATE: No geminate consonants. Oostendorp, ROA-84.
2. $\sigma \perp {}_H[$, etc. — *COMPLEX(T): A tone-bearer must not be associated with more than one tone. Myers, ROA-6.
3. $\sigma \perp {}_C[$ — NOCOMPLEXONSETORRHYME. Noske, ROA-109.
4. $\mu \perp {}_C[$ — *COMPLEX: No complex onset or coda. Kenstowicz, ROA-103.
5. $rime \perp {}_{nas}[$, etc. — RHYME HARMONY: All segments in the rhyme must share any nasal specification. Yip, ROA-81, ROA-135.

d. **Morphophonology**

1. $Red \perp {}_F[$, — RED = Foot. Carleton & Myers, ROA-16. (Also need ${}_{Red}[\rightarrow {}_F[$, $]_{Red} \rightarrow]_F$.)
2. $lab \perp {}_M[$ — MONOLOG: The edges of a morphological domain should be crisp; no feature should be linked both to an edge segment of that domain and to an element outside of the domain. Oosetndorp, ROA-84. (Also need $lab \perp]_M$.)
3. $(\ {}_{\mathtt{x}}[$ and ${}_V[\ ) \perp \underline{Root}$ — FTONSET$^{\{rt\}}$: Align(Ft that is in root, L, C or Root, L). Buckley, ROA-56.

7. Straitjacketing phonology keeps it local

Two important differences between ${}_F[\rightarrow {}_{PrWd}[$ and ALIGN(F, L, $PrWd$, L):

- The $\rightarrow$ family *doesn't measure distance.*
  E.g., $(\acute{\sigma}\sigma)(\acute{\sigma}\sigma)(\acute{\sigma}\sigma)\sigma$ violates ${}_F[\rightarrow {}_{PrWd}[$ twice, once for each non-initial foot.
- The $\rightarrow$ family *isn't only used for edges.*

Interestingly, Zoll (1996:137–38) has independently argued that licensing has just those properties (leading to her constraint COINCIDE(X,Y)):

> "There are two properties of licensing which distinguish it from the cases of affixation discussed [in M&P (1993)].
>
> "First, licensing of marked structure never involves an injunction to be as close to a strong position as possible. Rather, licensing *always* constitutes an all-or-nothing proposition whereby marked structures are licit in licensed positions but ill-formed everywhere else."
>
> "The second important difference is that licensing does not strictly involve coincidence of edges or distance from an edge, but is concerned rather with membership in a constituent which may be peripheral ... [e.g.] heavy syllables belong to the first foot."

Q: Is this local version of alignment powerful enough?
A: Perhaps so. For cases where it's really necessary to measure distance, for example to control the width of a feature domain:

(33) a. $\sigma \rightarrow XDom$: X-domain should be as wide as possible (contain many σ's).
b. $\sigma \perp XDom$: X-domain should be as narrow as possible (contain few σ's).

Note that this trick, unlike GA, automatically specifies the units of measurement. It also avoids other definitional problems with GA.

Q: Is Generalized Alignment too powerful?
A: Probably. It's a family of non-local constraints that do addition. That lets us express very non-local, unattested phenomena.

Example of unwarranted power: The GA constraint in (34) wants the floating tone to anchor as close to the <u>center of the word</u> as possible (subject to higher-ranked constraints).

(34) Notes:

1. ´ denotes tone, not stress.
2. The n^{th} column records the degree of misalignment of the n^{th} syllable, at least if GA measures this in syllables rather than segments (or moras: see Mester & Padgett (1993)).
3. Assume that high-ranked faithfulness constraints rule out other candidates. For example, as there is only one floating tone underlyingly, $\acute{\sigma}\acute{\sigma}\acute{\sigma}\acute{\sigma}\acute{\sigma}\acute{\sigma}\acute{\sigma}$ is ruled out by DEP(H).

$\sigma\sigma\sigma\sigma\sigma\sigma\sigma$ + [H]	ALIGN(σ, R, H, R)							
a. $\acute{\sigma}\sigma\sigma\sigma\sigma\sigma\sigma$	0	*	**	***	****	*****	******	= 21
b. $\sigma\acute{\sigma}\sigma\sigma\sigma\sigma\sigma$	*	0	*	**	***	****	*****	= 16
c. $\sigma\sigma\acute{\sigma}\sigma\sigma\sigma\sigma$	**	*	0	*	**	***	****	= 13
♡ d. $\sigma\sigma\sigma\acute{\sigma}\sigma\sigma\sigma$	***	**	*	0	*	**	***	= 12
e. $\sigma\sigma\sigma\sigma\acute{\sigma}\sigma\sigma$	****	***	**	*	0	*	**	= 13
f. $\sigma\sigma\sigma\sigma\sigma\acute{\sigma}\sigma$	*****	****	***	**	*	0	*	= 16
g. $\sigma\sigma\sigma\sigma\sigma\sigma\acute{\sigma}$	******	*****	****	***	**	*	0	= 22
		↑ violations contributed by 2nd syllable's misalignment						candidate's total viols

If there were *two* floating tones, they'd want to anchor at 1/4 and 3/4 of the way through the word.

This kind of non-local behavior via GA is unattested to my knowledge. It is also beyond the power of known computational OT methods, in particular the finite-state method of Ellison (1995) and the context-free method of Tesar (1996). The primitive constraints are provably incapable of producing such behavior.

8. In dire straits: What about non-local phenomena?

Since OTP uses only the primitive constraints of §4, it claims that *all phonology is* *local*.

Some apparently non-local phenomena can be reanalyzed:

- *Metrical stress.* Most non-local constraints in the literature concern metrical stress, which has received both local and non-local analyses in the past.
  - *Local:* Non-OT, iterative accounts (e.g., Prince 1983, Halle & Vergnaud 1987, Kager 1993, Hayes 1985, 1995).
  - *Non-local:* McCarthy & Prince (1993) propose using Generalized Alignment constraints to measure the distance from each foot to the edge of the word.
  - *Local:* Eisner (1997c) gives an OTP typology of metrical stress. See §9.
  - *Local:* "Incremental" constraints as described in §11. (Cf. Kager (1994), who argues for a greedy ALIGN evaluated "foot by foot.")
- *Intervocalic phenomena* (e.g., lenition). A constraint like *VsV (Green & Kenstowicz 1995) appears non-local, since [s] must look to *both* sides to decide whether it can surface as s or must become z. However, a local reanalysis is possible.

  Sample reanalysis: For *VsV, say that /s/ *always* wants to surface as [z], but only succeeds in the VsV context. For instance: (*cor* and *cont*) → *voi* rules out [s] in favor of [z]. It is outranked by $]_z \rightarrow (\ \underline{]_{voi}}$ or $_V[\ )$, which says that any surface [z] not underlyingly voiced is followed by a vowel, and also by the mirror image of this, so that such a [z] must also be preceded by a vowel. Here $]_z$ abbreviates $(\ ]_{cor}$ and $]_{cont}$ and $]_{voi}\ )$.

However, *reduplication* occupies a special role in phonology, in that it is inherently non-local; it cannot be reanalyzed as local.

Therefore, to handle reduplication in OTP we need a representational trick (similar to Clements 1985). Translate the Correspondence account of McCarthy & Prince (1995) into OTP as follows:

a. As for all relations, OTP can enforce Correspondence only locally, so Correspondent elements must always overlap on the timeline,
b. Thus, I-B faithfulness requires I and B to occupy the same portion of the timeline. (on separate input and output tiers)
c. B-R faithfulness apparently requires R and B to occupy the same portion of the timeline. But this would rule out B-R juncture effects. which require B to precede R or vice-versa. (e.g., enforcement of *VhV in Javanese)
d. So instead require R (on the output tier) and a copy of B (on its own special tier) to occupy the same portion of the timeline.
e. Gen produces only candidates in which this copy of B is perfect. Thus, Gen must know how to do reduplication of morphemes, not just affixation.
f. Now all the non-locality is handled within Gen (the locus of morphology: see §5). The violable constraints remain local.

(35) Some candidates produced by Gen on RED(*bədah*)-e. In Javanese, first candidate wins.

a.	bədah-e	Input tier (used for I-B faithfulness)
	bəda bəda -e	Output tier: passed to phonetics (here violates MAX-IO)
	[Red][Base][Af]	Morphemic tier: mentioned by some constraints
	bəda ↙	Exact copy of base (used for B-R correspondence)
b.	bədah-e	
	bədahbədah-e	Satisfies MAX-IO, but violates surface constraint **VhV*
	[Red][Base][Af]	
	bədah↙	Exact copy of this candidate's base (enforced by Gen)
c.	bədah-e	
	bədahbəda -e	Satisfies MAX-IO & **VhV*, but not DEP-BR, i.e., $C \rightarrow \underline{\underline{C}}$
	[Red][Base][Af]	
	bəda ↙	Exact copy of this candidate's base (enforced by Gen)
d.	bədah-e	
	bəda bədah-e	Satisfies MAX-IO, but not **VhV* or MAX-BR, i.e, $\underline{\underline{C}} \rightarrow C$
	[Red][Base][Af]	
	bədah↙	Exact copy of this candidate's base (enforced by Gen)

In a language also requiring I-R faithfulness (McCarthy & Prince's (1995) Full Model), Gen must put two copies on the input tier: bədah bədah-e.

Haplology is a related example that may also be intrinsically non-local. (Yip 1995)

9. Stress typology

Eisner (1997c) proposes a small set of primitive constraints, which are freely reranked to get the attested iambic systems. Replacing each constraint by its mirror image gives the attested trochaic systems. (All rankings have been tested exhaustively by computer.)

The result is a unified fine-grained account of the following phenomena described by Hayes (1995):

(36)
1. asymmetric foot shape typology
2. iambic lengthening
3. unbounded stress
4. simple word-initial and word-final stress
5. LR and RL footing, but no clear cases of RL iambs
6. syllable and foot extrametricality
7. no cases of final-syllable extrametricality for LR trochees **(new!)**
8. strong and weak prohibitions on degenerate feet
9. word-level stress, including prominence-based systems

The asymmetries in (36) are reduced to (i) the universal onset-coda asymmetry and (ii) the universal tendency of extrametricality to be final.

A few key ideas in this analysis:

(37) Alternating stress is the result of constraints against unary feet (which prevents stress clash) and against stress lapse.

(38) Stress prefers to fall on weak moras u_w, which carry weight.

- Consequence in iambic systems: The strong (right) edge of the foot likes to be supported by a weak mora, so stressed light syllables are avoided except when necessary to prevent lapses. This explains iambic foot form, iambic lengthening, and unbounded weight-prominence systems.
  (Whereas in trochaic systems, stressed lights are no worse than stressed heavies: there's never a weak mora at the strong (left) edge.)
- Consequence in trochaic systems: Moraic trochees of the form $(\acute{\mu}_s \mu_w)$ are avoided because the stress mark ´ would like to spread rightward onto μ_w. Such spreading yields syllabic trochees.
  (Whereas in iambic systems, stress starts out on μ_w and has no incentive to spread leftward: so there are no syllabic iambs.)

(39) The "natural" lapse-avoiding pattern on an odd string of light syllables is $\sigma\acute{\sigma}\sigma\acute{\sigma}\sigma$, which is LR iambs $(\sigma\acute{\sigma})(\sigma\acute{\sigma})\sigma$ or RL trochees $\sigma(\acute{\sigma}\sigma)(\acute{\sigma}\sigma)$. The other common case, LR trochees, results from right extrametricality, $(\acute{\sigma}\sigma)(\acute{\sigma}\sigma)\langle\sigma\rangle$, when right extrametricality is outranked by a desire to include another stress when possible: $(\acute{\sigma}\sigma)(\acute{\sigma}\sigma)(\acute{\sigma}\sigma)$. RL iambs aren't attested because left extrametricality is extremely rare (Kashaya).

Novel prediction from (39): When extrametricality is high-ranked enough to be surface-true, LR trochees are impossible. These two properties should be in *complementary distribution*: they are just different manifestations of the same mechanism (namely, the right edge of the word pushing feet away).

Confirmation: Among trochaic languages, Hayes (1995) lists 32 that are LR and 21 that have final-syllable extrametricality. There is no overlap! That is, no language has preantipenultimate stress on even strings $(\acute{\sigma}\sigma)(\acute{\sigma}\sigma)\sigma\langle\sigma\rangle$, but not on odd strings $(\acute{\sigma}\sigma)(\acute{\sigma}\sigma)\langle\sigma\rangle$. The right edge of the word can push feet away, but since primitive constraints are local, it can't demand that they be pushed all the way to the left of the word.

10. Computational issues: Theories in straitjackets are docile

Q: Gen produces infinitely many candidates. How do we find the best?
A: By using *intensional descriptions* of the infinite sets. For example, $son \rightarrow voi \gg \mu_w \perp voi$ yields "Utterances in which obstruent codas are voiceless and sonorants are voiced."

If we stick to the primitive constraints, we can use finite-state automata as our intensional descriptions. E.g., the infinite set of candidates that survive constraints 1–5 can be described in finite space with an automaton. Then we use constraint 6 to narrow this set down further, etc.
(Strategy is due to Ellison (1994); Eisner (1997b) gives an efficient version.)

Analogy: In mathematics, we don't work directly with the infinite sum

$$\frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \frac{1}{4\cdot 5} + \ldots$$

because that would take forever. Instead we manipulate the notation $\sum_{i=1}^{\infty} \frac{1}{i\cdot(i+1)}$. This lets us draw interesting conclusions without processing the terms one by one:

$$\sum_{i=1}^{\infty} \frac{1}{i\cdot(i+1)} = \sum_{i=1}^{\infty} \frac{1}{i} - \frac{1}{i+1} = \sum_{i=1}^{\infty} \frac{1}{i} - \sum_{i=1}^{\infty} \frac{1}{i+1} = (1 + \sum_{i=2}^{\infty} \frac{1}{i}) - \sum_{i=2}^{\infty} \frac{1}{i} = 1$$

BUT: To find the optimal candidate is NP-hard on the size of the grammar (Eisner 1997b). So while the automaton algorithm above is usually efficient, *any* algorithm will be slow for a pathological grammar. This is unfortunate for learning theories that may blunder into such a grammar and try to test it.

In addition to the algorithm to find the optimal candidate, we can also characterize the expressive power of OTP:

(40) a. *Equal* in power to OTFS (Ellison 1994), in which Gen is a finite-state transducer and the constraints are arbitrary weighted FSAs. Any formal OTP grammar can be converted to a formal OTFS grammar, and vice-versa.
OTP is nonetheless more restrictive. OTP grammars are more fine-grained, so they make stronger predictions about the effect of reranking constraints. Also, they are limited to a smallish set of universal tiers.

b. *Less* power than if Generalized Alignment were allowed. The crucial example is (34)—a funny trick that GA can do but OTP can't.

c. *More* power than systems of ordered rewrite rules, i.e., finite-state transducers. The crucial example is a trick that OTP can do but finite-state transducers can't (adapted from a non-linguistic example in Frank & Satta 1998):

This grammar puts H tones on either the high *or* the low vowels—whichever are fewer. ${}_D[\rightarrow {}_{PrWd}[$, $]_D \rightarrow]_{PrWd}$, $H \perp]_V$, $hi \rightarrow (H$ or $D)$, $(lo$ and $D) \rightarrow H \gg H \perp H$

11. A possible extension: Incremental evaluation of constraints

The example in (40c) should make us uncomfortable about OT! Linguistically, grammars that count like that are unattested. So OT appears too powerful.

Indeed, ordered rewrite rules were always *descriptively* adequate. OT was supposed to give us more *elegant or explanatory* descriptions, not admit counting tricks as a possibility in human language. So:

(41)

Can we somehow pare OT back to the same descriptive power as ordered rewrite rules—the power of a finite-state transducer?

There are also computational reasons to ask this question. Transducers are efficient, well understood, and invertible. In particular, when an OT grammar can be compiled into a transducer, we immediately obtain wonderful things:

(42) a. a much faster generation algorithm (UR $\rightarrow$ its optimal SR or SRs)
b. a comprehension algorithm (SR $\rightarrow$ the UR or URs for which it's optimal)

Note that (42b) (which would be just as fast as (42a)) produces all *possible* URs; this possibly large set can be efficiently

(43) a. intersected with a lexicon to find legitimate URs
b. intersected with lexical constraints to guess URs for novel words
c. used for phonological learning

Well, imagine that we allow different *modes of evaluation* for each primitive constraint:

(44) *Traditional Summing*: The candidates with the fewest total violations survive to be considered by subsequent constraints.

(45) *k-Bounded*: Like (44), but the constraint can only distinguish among 0, 1, 2, ..., $k-1$, or "k or more" violations; it can't count beyond k.
(For example, 1-bounded evaluation just asks "were there violations?" without asking how many.)

Frank & Satta (1998) and Karttunen (1998) recommend allowing only k-bounded constraints, since then we remain within finite-state power. This is true, but somewhat awkward as a solution:

- *Big inelegant transducers:* A grammar that sets k just large enough to handle "supercalifragilisticexpialidocious" would yield a much larger and more redundant transducer than would the comparable derivational grammar.
- *Lack of generalization:* Moreover, unlike the derivational grammar, it would fail to generalize to longer words and phrases.

So I'll propose two additional modes:

(46) *Incremental LR*:
- The constraint scores each candidate incrementally from Left to Right.
- This LR scoring proceeds *in parallel* for all candidates, staying in lockstep with reference to the candidates' common UR.
- Sudden death for any candidate as soon as it incurs a violation, unless all other surviving candidates simultaneously suffer an equally bad violation.

(47) *Incremental RL*: The mirror image of (46).

Summing evaluation is shown in (48–49). To understand incremental evaluation, consider a case where all (remaining) candidates have 4 syllables. Then (50) operates *as if* exploded into (51). It prefers to *postpone the pain* of NoCoda violations as long as possible, even at the cost of having more violations later.

(48) NoCoda: $]_\sigma \perp]_C$ / traditional summing

(49)

		NoCoda
♡	bantondibo	**
	bantodimbon	***

(50) NoCodaLR: $]_\sigma \perp]_C$ / incremental LR

(51)

		NoCoda-σ_1	NoCoda-σ_2	NoCoda-σ_3	NoCoda-σ_4
	bantondibo	*	*!		
♡	bantodimbon	*		*	*

(52) NoCodaRL: $]_\sigma \perp]_C$ / incremental RL

(53)

		NoCoda-σ_4	NoCoda-σ_3	NoCoda-σ_2	NoCoda-σ_1
♡	bantondibo			*	*
	bantodimbon	*!	*		

These ideas can be formalized, and the following theorem holds:

(54) *Theorem:* Suppose an OT grammar consists of primitive constraints each of which is evaluated as LR, RL, or k-bounded. (That is, the grammar never employs the traditional summation of (44).) Then one can construct a finite-state transducer that is equivalent to the grammar.

Now the question is: Can we get away with this linguistically?

Tentatively, yes:

Most of the time, incremental evaluation is indistinguishable from summing evaluation. When does it matter? Consider (51):

- Higher-ranked constraints have forced us to choose between satisfying NoCoda on the 2nd vs. the 3rd syllable: we can't satisfy it on both.
- This corresponds to the crucial-ranking pattern *! | * in the tableau.)
- But the situation as depicted in (51) doesn't tend to arise, since the two codas don't interact in any way.
- Such tradeoffs arise (for syllabification) only in the context of directional syllabification (see Mester & Padgett 1994)—which *is* resolved LR or RL!

Indeed, such forced tradeoffs—where only one constraint is at issue, but the language must choose *where* to violate it—are generally resolved by violating as late (LR) or as early (RL) as the higher-ranked constraints will allow. Examples:

(55)
1. Directional syllabification: /cvcccv/ → [cv.cVc.cv] vs. [cvc.cV.cv]. To favor the latter, as in Cairene Arabic, evaluate **Dep** incrementally LR so as to postpone epenthetic material; this is classically known as LR syllabification.
2. Footing: $(\sigma\sigma)(\sigma\sigma)\sigma$ vs. $(\sigma\sigma)\sigma(\sigma\sigma)$ vs. $\sigma(\sigma\sigma)(\sigma\sigma)$. (PARSE($\sigma$): $\sigma \rightarrow F$.)
3. Infixation: Tagalog [*gr-um-adwet*]. (NOOVERLAP: $]_{Affix} \perp Stem$.)
4. Docking of floating features (e.g., tone). (Either $V \rightarrow H$ or $V \perp H$.)
5. Resolution of OCP violations (e.g., Grassman's Law). (Likewise.)
6. End rule: $\grave{\sigma}\sigma\grave{\sigma}\sigma\acute{\sigma}$. (Likewise.)
7. Marking of special domains that license additional material, like the first foot in a word. (Likewise.)

Warning: To evaluate a form incrementally, a constraint must be defined so as to specify not just how many violations there are, but also *where* they fall. I have not yet addressed this issue for the primitive constraint families.

12. A possible extension: Capturing tier adjacency

In a standard autosegmental representation (Goldsmith 1990), even feet with gaps between them would be treated as adjacent on their tier, as would adjacent tones and adjacent vowels. This makes it easy to discuss OCP effects, long-distance spreading, tonology, the notion of "leftmost X," etc.

The strictly local timeline representations of (15b) lose this adjacency. On the timeline, two H tones on the same tier can't see each other through intervening consonants.

One way out is provided by Optimal Domains Theory (Cole & Kisseberth 1994): each phonetic feature silently projects a wider domain, and adjacency is defined locally in terms of these domains. This is functional but somewhat clumsy.

A possible alternative, more in the spirit of autosegmental phonology:

(56)
- Every constraint has a set of relevant tiers.
- The relevant tiers are those that participate directly in the constraint, plus any others mentioned on the side.
- When evaluating a constraint, we **collapse** (ignore, skip over) any time intervals where nothing is happening on the relevant tiers.

Under this "collapsing convention," we could write the OCP for H tone as in (57). This says that adjacent H's must be separated by a $]_{PrWd}$—where two H's are considered to be adjacent if there are no H's, PrWd's, or L's between them (these being the relevant tiers mentioned in the constraint). Effectively, (57) treats (58) as if it were (59).

(57) OCP(H): $(\ ]_H$ and $_H[\ ) \rightarrow {}_{PrWd}[$ (L)


```
(58) [              PrWd            ][PrWd ]
        [ H ]  [ H ]           [ H ]  [ H ]
                       [ L ]
      b   u   d   u   g   o   t   i   n   o
               . . . other tiers . . .

(59) [    PrWd    ][ PrWd ]
     [ H ][ H ]  [ H ][  H   ]
          [ L ]
```

Similarly, under this convention, (60) requires that any H tone be spread onto the vowel to its right. (61) restricts this requirement to within the PrWd (just as in (57)).

(60) HHARM(Right): $_H[\ \perp\ ]_V$
(61) HHARMBOUNDED(Right): $(\ ]_H$ and $_V[\ )\ \perp\ ]_{PrWd}$

Note that (62) no longer enforces right extrametricality: it mentions only F and PrWd as relevant tiers, causing (64) to be interpreted as (65). Instead we need to use (63), which recognizes syllabified material and requires it to intervene.

(62) EXTRAMETRICALITY: $]_F \perp]_{PrWd}$
(63) EXTRAMETRICALITY: $]_F \perp]_{PrWd}$ (σ)

```
(64) [      PrWd       ]
     [  F  ][  F  ]
     [ σ ][ σ ][ σ ][ σ ][ σ ]
     . . . other tiers . . .

(65) [  PrWd  ]
     [ F ][ F ]
```

A formal version of the collapsing convention:

(66) a. Any two brackets are considered to coincide in time unless they are separated by an entire constituent $_x[\quad]_x$ on a relevant tier, or more generally by a configuration $_x[\quad]_y$ where x and y are both relevant tiers.
b. That is, configurations $_x[\quad]_y$ on the relevant tiers *resist* collapse.
c. Equivalently, imagine conflating the relevant tiers to get a sequence of labeled brackets in time. Any contiguous subsequence of the form
$\cdots]\ \leq]\ \leq]\ \leq [\leq\ [\leq\ [\leq\ [\cdots$
(i.e., 0 or more] followed by 0 or more [)
is considered to be contemporaneous for purposes of this constraint.

Really this interprets adjacency as in Goldsmith (1990), while allowing each constraint to specify which tiers to conflate. (Tier conflation is traditionally an operation that happens at some point during a derivation—but OT is not derivational.)

A weaker alternative to (66c) is also worth considering. This would not collapse (64) into (65):

(67) Consider only sequences of the form
$\cdots] =] =] \leq [= [= [= [\cdots$
to be contemporaneous.

The collapsing convention of (66) or (67) can easily be used to solve the problem noted in (24d).

It generalizes the previous solution to (24d) (outlined in Albro 1997), which effectively combined (67) with the notion that all surface tiers were relevant to a constraint that mentioned any surface tier, and all underlying tiers were relevant to a constraint that mentioned any underlying tier.

13. What role do these primitive constraints play in OT?

Three kinds of constraints:

- Primitive: the implication and clash families.
- Compound: Expressible as a monolithic block of primitive constraints in fixed order. (Kennedy (1996) uses blocks of Align constraints.)
- Complex: Any constraint not expressible in this restricted framework.

The balance among these remains to be seen. It is not yet clear what compound or complex constraints are actually needed (and which of the primitive constraints are *not* needed!).

We must also discover which of the formally possible primitive constraints are favored in real languages (on phonetic or other grounds), and what rankings are favored. OTP claims that languages use only local constraints; but it does not say *which* local constraints.

Meanwhile,

- Primitive constraints are "safe to use." They're simple, radically local, and ubiquitous.
- The restricted version of OT allowing *only* primitive constraints—called OTP—is easy to reason within and is computationally tractable.
- OTP is the simplest explanation that stands a chance. Let's refine it against the data, adding new core constraints only as we're forced to.
- If OTP is close to correct, it may be fruitful to reanalyze languages and typologies within OTP. (E.g., Eisner (1997c) gives a detailed reanalysis of stress typology that has some empirical benefits.)

References (exclusive of §6)

Albro, Daniel M. 1997. *Evaluation, Implementation, and Extension of Primitive Optimality Theory.* M.A. thesis, UCLA.
Albro, Daniel M. 1998. Three formal extensions to Primitive Optimality Theory. In Mark Ellison (ed.), Proceedings of the Fourth Meeting of the ACL Special Interest Group in Computational Phonology. Association for Computation Linguistics. Quebec, July.
Bird, Steven, & Ewan Klein. 1990. Phonological events. *Journal of Linguistics* **26**, 33–56.
Buckley, Eugene. 1995. Alignment and constraint domains in Manam stress. Ms. ROA-56.

Clements, G. N. 1985. The problem of transfer in nonlinear morphology. Cornell Working Papers, Cornell University, Ithaca, NY, Fall.
Cole, Jennifer, & Charles Kisseberth. 1994. An optimal domains theory of harmony. *Studies in the Linguistic Sciences* **24**: 2.
Crowhurst, Megan 1994. Prosodic alignment and misalignment in Diyari, Dyirbal, and Gooniyandi: an optimizing approach. *WCCFL* **13**. ROA-19.
Eisner, Jason M. 1997a. What constraints should OT allow? LSA Annual Meeting, Chicago, January.
Eisner, Jason. 1997b. Efficient generation in primitive Optimality Theory. Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Association for Computational Linguistics, Madrid, July.
Eisner, Jason. 1997c. FootForm decomposed: Using primitive constraints in OT. SCIL 8. MIT Working Papers in Linguistics, Cambridge, MA.
Ellison, T. Mark. 1994. Phonological derivation in optimality theory. Proceedings of **COLING**, 1007-1013.
Féry, Caroline. 1994. Umlaut and Inflection in German. Ms. ROA=34.
Frank, Robert, and Giorgio Satta. 1998. Optimality theory and the generative complexity of constraint violability. *Computational Linguistics* 24(2):307–315.
Goldsmith, John A. 1976. *Autosegmental Phonology.* Ph.D. thesis, MIT.
Goldsmith, John A. 1990. *Autosegmental and Metrical Phonology.* Cambridge, MA: Blackwell.
Green, Thomas and Michael Kenstowicz. 1995. The Lapse constraint. **FLSM** 6,1: 1-14. Bloomington, Ind: Indiana University Linguistics Club. ROA-101.
Green, Thomas. 1994. The conspiracy of completeness. Proceedings of Rutgers Optimality Workshop I. ROA-8.
Halle, Morris and Jean-Roger Vergnaud. 1978. Metrical structures in phonology. Ms., Dept. of Linguistics, MIT.
Hammond, Michael. 1995. Syllable parsing in English and French. Ms. ROA-58.
Hayes, Bruce. 1985. Iambic and trochaic rhythm in stress rules. *BLS* **11**, 429–446.
Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and Case Studies.* University of Chicago Press.
Itô, Junko, & Armin Mester. 1994. Realignment. Proceedings of the June 1994 Utrecht Prosodic Morphology Workshop.
Kager, René. 1993. Alternatives to the Iambic-Trochaic Law. *LI* **11**, 381–432.
Kager, René. 1994. Ternary rhythm in alignment theory. Ms. ROA-35.
Karttunen, Lauri. 1998. The Proper Treatment of Optimality in computational phonology. Proceedings of FSMNLP'98 (International Workshop on Finite-State Methods in Natural Language Processing), 1–12. Bilkent University, Ankara, Turkey.
Kirchner, Robert. 1993. Turkish vowel harmony and disharmony: an optimality theoretic account. Proceedings of Rutgers Optimality Workshop I. ROA-4.
McCarthy, John, & Alan Prince. 1986. Prosodic morphology. Ms., Brandeis University.
McCarthy, John, & Alan Prince. 1993. Generalized alignment. *Yearbook of Morphology*, ed. Geert Booij & Jaap van Marle, pp. 79-153. Kluwer.
McCarthy, John and Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill Beckman et al., eds., Papers in Optimality Theory. UMass, Amherst: GLSA. 259–384.
Mester, Armin and Jaye Padgett. 1994. Directional syllabification in Generalized Alignment. Phonology at Santa Cruz 3, October. ROA-1.
Myers, Scott. 1994. OCP effects in Optimality Theory. Ms. ROA-6.
Polgardi, Krisztina. 1995. Derived Environment Effects and Optimality Theory. Handout, Tilburg "Derivational Residue" Conference. ROA-93i.
Prince, Alan, & Paul Smolensky. 1993. *Optimality theory: constraint interaction in generative grammar.* Technical Reports of the Rutgers University Center for Cognitive Science.

Prince, Alan. 1983. Relating to the Grid. *LI* **14**, 19–100.

Selkirk, Elizabeth. 1980. Prosodic domains in phonology: Sanskrit revisited. In Mark Aranoff and Mary-Louise Kean, eds., *Juncture*, pp. 107–129. Anna Libri, Saratoga, CA.

Smolensky, Paul. 1995. On the structure of the constraint component Con of UG. Talk at UCLA, April 7. ROA-86.

Tranel, Bernard. 1994. French liaison and elision revisited: A unified account within Optimality Theory. In Claudia Parodi *et al.*, Aspects of Romance Linguistics: Selected Papers from the Linguistic Symposium on Romance Languages XXIV. Washington, D.C.: Georgetown University Press. ROA-15.

Yip, Moira. 1994. Phonological constraints, optimality, and phonetic realization in Cantonese. Ms. ROA-14.

Yip, Moira. 1995. Identity Avoidance in Phonology and Morphology. Proceedings of the Conference on Morphology and its relation to Syntax and Phonology. University of California at Davis, May. ROA-82.

Zec, Draga. 1988. *Sonority constraints on prosodic structure*. Ph.D. dissertation, Stanford University.

Zoll, Cheryl. 1996. *Parsing Below the Segment in a Constraint Based Framework*. Ph.D. dissertation, University of California at Berkeley. ROA-143.