

# Remarks and Replies

## On Scrambling: A Reply to Bošković and Takahashi

*John Frederick Bailyn*

In this article I argue against Bošković and Takahashi's (1998) analysis of scrambling as base generation (with lowering for  $\bar{A}$ -cases). I present evidence from Russian of scope and antireconstruction effects and scrambling/*wh*-movement parallels, all implicating a "classical" overt movement account of  $\bar{A}$ -scrambling. I then discuss theoretical issues unresolved by the base generation/lowering account. Having shown that  $\bar{A}$ -scrambling is (upward) movement, I argue that the account of  $A$ -scrambled arguments as base-generated also loses its force. In conclusion I suggest an alternative way to eliminate the apparent optionality associated with scrambling, while maintaining the classical analysis of scrambling as upward movement.

*Keywords:* scrambling, reconstruction, optionality,  $\bar{A}$ -movement, constraints on movement, word order, Russian syntax

Bošković and Takahashi (1998) (hereafter B&T) claim that scrambled elements are base-generated in their surface (adjunct) position and undergo a postsyntactic process of lowering into  $\theta$ -position, in  $\bar{A}$ -cases, under a view where  $\theta$ -roles are formal features, in the sense of Chomsky 1995. By eliminating scrambling itself, replacing it with base generation, this approach eliminates the optionality associated with scrambling, reversing the standard analysis that scrambling is optional, semantically vacuous movement (Ross 1967, Saito 1989, 1992).<sup>1</sup>

However, as I will show, B&T's approach runs into various difficulties: it leaves restrictions on scrambling unexplained; makes certain predictions that are falsified by the data; loses significant generalizations; and does not fully solve the problem that it sets out to address, namely, that of the apparent optionality of scrambling. Given these difficulties, I argue that the "classical" view of scrambling as upward movement must be maintained.

The article is structured as follows. Section 1 presents the classical view of scrambling and

I would like to thank Olga Arnaudova, Leonard Babby, John Bowers, Barbara Citko, Marcel den Dikken, Daniel Finer, Richard Larson, James Lavine, Gil Rappaport, Mark Volpe, John Whitman; audiences at WCCFL XIII, the Arizona International Workshop on Word Order and Scrambling, and at Princeton, Leipzig, and Potsdam Universities; and anonymous *LJ* reviewers for helpful discussion. All mistakes remain my own. This research was partly supported by IREX and the Social Science Research Council.

<sup>1</sup> The following works propose movement accounts of scrambling: for German, see Webelhuth 1989; for Japanese, see Saito 1989, 1992, Miyagawa 1997, forthcoming; for Russian, see Yadroff 1992, Müller and Sternefeld 1993, Bailyn 1995, forthcoming; for Hindi, see Mahajan 1990; for Korean, see Lee and Santorini 1994. For nonmovement accounts, see Bayer and Kornfilt 1994, Neeleman 1994.

B&T's alternative, identifying predictions that can be used to test them. Section 2 presents evidence from Russian against the predictions made by B&T's approach. Section 3 raises some theoretical issues that are problematic for their approach and returns to the original question of optionality. Section 4 discusses A-scrambling under B&T's account. An alternative way to eliminate the apparent optionality of scrambling while maintaining upward movement is presented in section 5.

## 1 Two Views of Scrambling

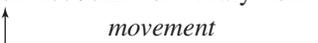
Let us begin by looking at the "classical" analysis of long-distance scrambling, then at B&T's alternative. The scrambling phenomenon is observed in sentence pairs like (1a–b) from Japanese (B&T's (3a–b)). Members of such pairs differ in word order, but are identical in grammatical relations, truth conditions, and morphology.

- (1) a. John-ga [Mary-ga *sono hon-o* katta to] omotteiru.  
 John-NOM Mary-NOM that book-ACC bought that thinks  
 'John thinks that Mary bought the book.'  
 b. *Sono hon-o* John-ga [Mary-ga *t* katta to] omotteiru.

Starting with Ross (1967), many researchers have taken pairs like (1a–b) to be related by a "scrambling" transformation that derives (1b) from (1a) by reordering constituents.<sup>2</sup> The contrast between the classical analysis and B&T's alternative concerns whether or not (1a) is the transformational source sentence for (1b). The classical view, which I argue for in this article, claims that it is; B&T's alternative argues that the instances of *sono hon-o* in (1a) and (1b) occupy distinct base-generated positions.

### 1.1 The Classical Analysis of Scrambling

The classical view takes the structural relationship between a predicate and its arguments and modifiers in scrambling languages to resemble that found in nonscrambling languages. Specifically, arguments and modifiers begin in a local relation with their associated predicates; only later are they scrambled away. On this view (1a) must be the source structure in the pair (1a–b) since only in (1a) does *sono hon-o* 'that book' stand in a local structural relation with its predicate *katta* 'bought'. Thus, (1b) must derive from (1a) by scrambling *sono hon-o* away from its base position through raising, as in (2).

- (2) *Sono hon-o* John-ga [Mary-ga *t* katta to] omotteiru.  
 that book-ACC John-NOM Mary-NOM bought that thinks  


<sup>2</sup> Ross's original account limited the process to the local clause and would not have accounted for long-distance cases such as (1b). However, the term *scrambling* continued to be used in this derivational sense throughout the Government-Binding Theory period (see, among many others, Saito 1992). Under minimalist assumptions, on the classical view (1a) and (1b) share the same numeration. I return to this important assumption in section 3.

Within Government-Binding Theory the generally accepted view was that scrambling is an instance of Move  $\alpha$ , constrained by universal principles such as Subjacency, the Empty Category Principle (ECP), and the Condition on Extraction Domain (CED). Debates have involved the kind of movement (A- or  $\bar{A}$ -) and the correct analysis of cases like (3a–b) (B&T's (6a–b)), where a scrambled element is interpreted in its base position, despite surface displacement.

- (3) a. *Nani-o<sub>i</sub>* John-ga [Mary-ga *e<sub>i</sub>* katta ka] sitteiru.  
 what-ACC John-NOM Mary-NOM bought Q knows  
 'John knows what Mary bought.'  
 b. [*Mary-ga nani-o katta to*]<sub>i</sub> John-ga [Bill-ga *e<sub>i</sub>* itta ka] sitteiru.  
 Mary-NOM what-ACC bought that John-NOM Bill-NOM said Q knows  
 'John knows what Bill said that Mary bought.'

In (3a) the scrambled *wh*-object takes embedded scope, despite being in the main clause. Similarly for (3b): there an entire CP containing a *wh*-phrase is scrambled, but the *wh*-phrase contained in it is interpreted in the lower position.

Under this approach, then, (long-distance) scrambling is a form of  $\bar{A}$ -movement that *reconstructs* at LF to base position, where interpretation occurs.<sup>3</sup> This approach is illustrated in (4).

- (4) *Sono hon-o* John-ga [Mary-ga *t* katta to] omotteiru.  
 that book-ACC John-NOM Mary-NOM bought that thinks
- 

Reconstruction effects can be viewed as preliminary evidence in favor of the raising/movement analysis of scrambling insofar as they appear parallel to reconstruction effects found in *wh*-movement and other instances of  $\bar{A}$ -movement. We will return to Russian *wh*-movement and scrambling parallels below.

### 1.2 Bošković and Takahashi's (1998) Analysis of Scrambling

B&T invert the classical picture of (1a–b), claiming that the position of *sono hon-o* 'that book' in (1b) is in fact base-generated. On their proposal (1a–b) are not derivationally related to each other; however, at LF the scrambled (for B&T: base-generated) object undergoes an obligatory lowering operation, which derives an LF structure essentially identical to (1a).

- (5) *Sono hon-o* John-ga [Mary-ga *e* katta to] omotteiru.  
 that book-ACC John-NOM Mary-NOM bought that thinks
- 

<sup>3</sup> In more recent frameworks (Chomsky 1995) movement and reconstruction are replaced by copying and deletion (the copy theory). Reconstruction is handled by deleting the higher copy, leaving the lower copy at LF for interpretation. Assuming that under the copy theory constraints on movement are still handled as applying to the relationship between copies of a single element (effectively, a condition on chains), the copy theory either would render licit all cases of

Thus, in place of raising and LF reconstruction, there is *only* LF lowering. The differences between the two views and relevant definitions are summarized in (6).

- (6) *Differences between classical (4) and B&T's (5) approaches to scrambling*
- a.
    - i. In (4) the derivation involves *raising*. In (5) it does not.
    - ii. In (4) the derivation involves *reconstruction*. In (5) it does not.
    - iii. In (5) the derivation involves *obligatory lowering*. In (4) it does not.
  - b.
    - i. *Raising*: (standard) upward movement, forming a chain, leaving a trace, subject to constraints (Subjacency, ECP, CED, etc.)<sup>4</sup>
    - ii. *Reconstruction*: LF movement of an element back to its base-generated spot (or, in the copy theory, deletion of the higher copy) for interpretive purposes
    - iii. *Obligatory lowering*: feature-driven downward movement (as proposed by B&T)

The motivation for this proposal arises from the particular theory B&T assume, and from specific interpretive properties of scrambling. B&T frame their account within the Minimalist Program (Chomsky 1995), according to which there should be no optional movement, for reasons of economy. Although I differ with B&T on the driving force of scrambling and its mechanics, I share their determination to eliminate pure optionality from the grammar of natural language. This shared assumption serves as the backdrop for the discussion to follow.<sup>5</sup> As B&T observe, classical scrambling, on its movement analysis, appears to be optional. It is not motivated by Case considerations or the kind of feature checking that characterizes *wh*-movement.

Furthermore, scrambling appears to be inert semantically. That is, it appears to *obligatorily* reconstruct to its base position in the LF component (Saito 1989, 1992). In (3a–b), for instance, the displaced *wh*-phrase and *wh*-clause *must* be interpreted in their source site. With *wh*-movement and topicalization, by contrast, “undoing” the moved phrase is impossible; the movement is morphologically driven. B&T observe that “the fact that scrambling can be undone is puzzling under the standard assumption that, like *wh*-movement and topicalization, which cannot be undone, scrambling in the constructions under consideration involves overt  $\bar{A}$ -movement” (p. 354).<sup>6</sup>

---

lowering (an unwelcome result) or would remain indistinct from a theory with traces, which is what both B&T and I assume. In this article I will use the movement terminology (*base position*, *traces*, etc.) proposed in classical accounts and maintained by B&T.

<sup>4</sup> This use of the term *raising* should not be confused with the narrower use associated with NP-movement out of the complement of *seem*-type verbs, which is not relevant to this article. I use *raising* simply to implicate *upward* movement, in contrast to the lowering proposed by B&T.

<sup>5</sup> Various attempts have been made to soften the absolute ban on optionality. For example, Chomsky (1995) argues that optionality is possible only when two convergent derivations involve an identical number of steps, and Saito and Fukui (1998) argue that a language's basic structure-building mechanisms allow for certain optional movements. I share with B&T the belief that a theory admitting true optionality is weaker than one without optionality. I return to the central issue of the apparent optionality of scrambling in section 3.

<sup>6</sup> B&T provide examples showing that neither *wh*-movement nor topicalization allows scope dislocations. However, the claim that *wh*-movement and topicalization are never “undone” appears to be too strong. Standard reconstruction effects hold of these constructions, as shown in (i)–(iii).

B&T derive this apparent nonparallelism between scrambling and *wh*-movement/topicalization using an analysis in which the former is something quite different from movement. Their view relies crucially on the assumption that  $\theta$ -roles are formal features, which are parameterized in strength, like other formal features. A scrambling language (like Japanese) has weak  $\theta$ -roles and arguments need not merge into  $\theta$ -position; rather, they can be base-generated in scrambled position (adjoined to [Spec, IP]), as in (5), and lower into  $\theta$ -position at LF. This LF movement is, of course, obligatory (if it were not, the  $\theta$ -features would never be checked), and the Last Resort problem goes away. Japanese and English then differ in the strength of their formal  $\theta$ -features.<sup>7</sup>

Such an analysis accounts for the following generalizations: (a) Scrambling obligatorily reconstructs. (This follows directly for B&T—scrambled arguments always lower into  $\theta$ -position.) (b) Extraction out of scrambled elements is acceptable, as shown in (7).

- (7) a. John-ga [Mary-ga *sono hon-o* katta to] omotteiru.  
 John-NOM Mary-NOM that book-ACC bought that thinks  
 ‘John thinks that Mary bought the book.’
- b. *Sono hon-o<sub>i</sub>* [John-ga [<sub>CP</sub>[<sub>IP</sub>[<sub>CP</sub> Mary-ga *e<sub>i</sub>* katta to]<sub>k</sub> [Bill-ga *e<sub>k</sub>* itta]  
 that book-ACC John-NOM Mary-NOM bought that Bill-NOM said  
 to] omotteiru]].  
 that thinks  
 ‘That book<sub>i</sub>, John thinks that [that Mary bought *e<sub>i</sub>*]<sub>k</sub>, Bill said *e<sub>k</sub>*.’

(7a) involves no scrambling. In (7b) the direct object appears outside its containing CP, which is also not in its selected position, causing no violation. (This is expected under B&T’s account since after LF lowering, there is no violation.) (c) Scrambling contributes nothing to the semantics of the sentence. (This follows under B&T’s account on the assumption that interpretive relations are established at LF—the displaced element is always in  $\theta$ -position at LF.)

Given the assumptions shared in B&T 1998 and this article, B&T’s approach makes the following predictions:

- (8) a. *Prediction A*  
 There should be no interpretive effects associated with surface (scrambled) position.  
 (‘radical reconstruction’)

- 
- (i) [*That picture of himself<sub>i</sub>*]<sub>k</sub>, I know John<sub>i</sub> likes *t<sub>k</sub>*. (OK only if reconstruction takes place: Principle A)  
 (ii) ??[*That story about him<sub>i</sub>*]<sub>k</sub>, I think John<sub>i</sub> heard *t<sub>k</sub>*. (marginal only if reconstruction takes place: Principle B)  
 (iii) \*[*That story about John<sub>i</sub>*]<sub>k</sub>, I think he<sub>i</sub> heard *t<sub>k</sub>*. (\* only if reconstruction takes place: Principle C)

Clearly, any theory of reconstruction must accommodate (i)–(iii) on the one hand and the contrast between scrambling and topicalization on the other. I return to this issue in section 2.

<sup>7</sup>B&T’s account allows for another potential formalization of the typological difference between scrambling and nonscrambling languages, namely, the availability of a base-generated IP-adjunction site in the former but not in the latter. They themselves opt for the stronger claim that the strength of formal  $\theta$ -features is parameterized, and I therefore follow the stronger line in discussing their account.

b. *Prediction B*

There should be no constraints on the configurational relationship between the surface and  $\theta$ -positions of scrambled elements.

c. *Prediction C*

Only elements with  $\theta$ -roles should participate in scrambling.

Certain points must be clarified to see how these predictions follow from B&T's approach. Prediction A follows from the obligatory nature of LF lowering for B&T. Scrambled elements are in  $\theta$ -position at LF; they must therefore be interpreted there. This follows from the minimalist assumption that LF is the unique level of interface with the interpretive component of the grammar and therefore the only level at which interpretive information can be encoded.<sup>8</sup> Prediction B follows from the fact that lowering requires the absence of any trace in the base-generated position (B&T's claim); if there were such a trace, the Proper Binding Condition would be violated. This claim is central if B&T's approach is to allow the Proper Binding Condition to still rule out *overt* lowering, which remains generally unattested. "[O]vert lowering and lowering of elements that for independent reasons must leave traces are disallowed" (p. 351). Although B&T do not provide a specific formulation, the Proper Binding Condition ruling out overt lowering can be defined as in (9).

(9) *Proper Binding Condition* (applies at LF)

Traces must be bound.

Assuming standard c-command definitions of binding, (9) rules out any lowering process that leaves a trace, and it would rule out B&T's proposed lowering, were such movement to leave a trace. B&T's account specifically requires, therefore, that the lowering involved in (1b) *leave no trace*. "We follow Lasnik and Saito (1992) in assuming that movement does not have to leave a trace when no principle requires it. Then, the LF movement deriving [(1b)] does not have to leave a trace, rendering the Proper Binding Condition inapplicable" (p. 351). Importantly, on the assumption that a chain is defined as the relation between a trace and its antecedent (or between two copies of an element under the copy theory), in B&T's account the lack of a trace in turn denies the existence of any chain relating the base-generated site and the (LF)  $\theta$ -position. Constraints on chain formation therefore should not apply in such constructions, and scrambling should be generally unconstrained (prediction B).<sup>9</sup> Prediction C is a central prediction of B&T's theory, namely,

<sup>8</sup> An *LI* reviewer suggests that B&T's approach can allow for several different kinds of processes at LF; but they do not propose this additional complication of the grammar, and it would require further justification.

<sup>9</sup> The preceding remarks apply primarily to representational accounts such as B&T's and the classical alternative. A derivational account employing lowering might be better equipped to capture constraints on scrambling through economy conditions (Shortest Move) but would still face the problem of disallowing overt lowering while allowing scrambling (as lowering) yet at the same time accounting for its constrained behavior. Furthermore, a lowering process such as that proposed by B&T is impossible under a purely derivational theory such as Epstein et al.'s (1998), under which a term can only be attracted by another c-commanding term, ruling out lowering of the kind B&T suggest.

that scrambling is a property of  $\theta$ -marked elements.<sup>10</sup> We will see that these predictions are disconfirmed by data from Russian.

Russian is clearly a scrambling language in the original ‘stylistic’ sense intended by Ross (1967) and picked up by Saito (1989, 1992). It also has overt *wh*-movement, as discussed by Müller and Sternefeld (1993) and B&T themselves, among many others.<sup>11</sup> Russian scrambling is therefore a good testing ground for B&T’s account.<sup>12</sup>

## 2 Empirical Arguments against Scrambling as Lowering

### 2.1 Interpretive Effects of Overt Scrambling

Prediction A of B&T’s analysis is that scrambled position is not relevant to LF interpretation (for  $\bar{A}$ -scrambling). This follows from obligatory lowering, which expressly disallows any trace or copy in the scrambled position (the base-generated position for B&T). However, the prediction that no interpretive effects are associated with the scrambled position is too strong; we will now look at various cases in which scrambled position is directly relevant to interpretation.

*2.1.1 Russian Scope Effects* B&T do not state explicitly how scope is to be determined, which (as an *LI* reviewer points out) remains a serious unanswered question for the lowering account. However, B&T do not claim to be assuming anything other than the usual account of scope (namely, configurational LF relations), and thus scrambled orders should not produce changes in scope relations, since under B&T’s account scrambled elements undergo obligatory lowering at

<sup>10</sup> An *LI* reviewer points out that this prediction does not appear to follow directly from B&T’s analysis. The reviewer cites B&T’s example (13), repeated here as (i), in which an adjunct *wh*-phrase is scrambled in Japanese, as a possible counterexample to prediction C (addressed by B&T exactly because of its potentially unexpected behavior under their account).

(i) ?Naze Mary-ga [John-ga sono setu-o sinziteiru ka] sitteiru.  
 why Mary-NOM John-NOM that theory-ACC believes Q knows  
 ‘Mary knows why John believes in that theory.’

However, in discussing this example, B&T acknowledge that without piggybacking on the *wh*-feature forcing movement in (i), the phrase in question would have no motivation for lowering if base-generated, since it is an adjunct, and not  $\theta$ -marked. More precisely, then, prediction C might read: Only elements with  $\theta$ -roles should participate in scrambling, unless they are associated with *independent features* that force movement. In that case the objection might continue that Russian adverbs, which scramble freely, might well be associated with independent features that force their lowering from scrambled position, leaving B&T’s base generation account intact. Recall, however, that B&T claim that the typological difference between scrambling and nonscrambling languages lies in the strength of  $\theta$ -features. Therefore, the effects of the parameter should not be found on any elements other than  $\theta$ -marked elements. Prediction C might then read: Only elements with  $\theta$ -roles should behave differently in scrambling languages and nonscrambling languages. (Thanks to Richard Larson for discussion of this issue.)

<sup>11</sup> Some recent accounts of Slavic *wh*-fronting have concluded that various apparent instances of *wh*-movement are in fact instances of scrambling or its discourse equivalent, focus movement. See Stepanov 1997, Bošković 1997, 1998, Boeckx and Stjepanović 2000, and Strakhov 2001. Such analyses rely crucially on the (classical) analysis of scrambling as upward movement; the resulting generalizations are lost if scrambling does not involve movement. Thus, it appears difficult to maintain both this type of analysis and B&T’s.

<sup>12</sup> Scrambling judgments vary considerably across Russian dialects. Scrambling is accepted freely in some (see, e.g., Yadroff 1994) and is far more restricted in others. The range of possibilities in itself argues for a more fine-grained notion of scrambling, one that is possible with stricter and milder constraints on overt raising, rather than a binary parameter of weak versus strong  $\theta$ -features.

LF into  $\theta$ -position to check  $\theta$ -features. At LF they therefore occupy canonical argument position. The scrambled position is simply not available as a position for scope interpretation. Prediction A of B&T's analysis (namely, that no interpretive effects should be associated with the surface position of scrambled elements) should thus apply to scrambling and scope in Russian. And yet, as the following contrast shows, scrambled orders do differ from nonscrambled orders with respect to scope in Russian.<sup>13</sup>

- (10) a. Kto-to            xočet, čtoby Boris uvidel každogo mal'čika.  
 someone-NOM wants that Boris saw [every boy]-ACC  
 'Someone wants Boris to see every boy.'  
 (i)  $\exists x \forall y$     (ii)  $*\forall y \exists x$
- b. [Každogo mal'čika]<sub>i</sub> kto-to            xočet, čtoby Boris uvidel  $t_i$ .  
 [every boy]-ACC someone-NOM wants that Boris saw  
 'Every boy someone wants Boris to see.'  
 (i)  $*\exists x \forall y$     (ii)  $\forall y \exists x$

(10a) and (10b) are both unambiguous, but with different relative scope. The lack of ambiguity in (10a) is expected under standard versions of Quantifier Raising (QR) as a clause-bounded process. The possibility of wide scope for *every boy* in (10b), however, goes against the prediction made by the lowering hypothesis. This fact is a problem for an account that requires obligatory lowering into  $\theta$ -position at the same level where scope relations are determined. Both sentences should have the same interpretation.<sup>14</sup>

2.1.2 *Russian Antireconstruction Effects* It is well known that reconstruction with *wh*-movement differs for different kinds of phrases, as shown in (11). (The same holds for English topicalization. See Huang 1993 and Heycock 1995 for discussion.)

- (11) a.  $*[How\ proud\ of\ John]_k$  do you think  $he_i$  should be  $t_k$ ?  
 b.  $[Which\ question\ that\ Gore_i\ got\ during\ the\ debate]_k$  do you think  $he_i$  messed up on  $t_k$  the worst?

In (11a) reconstruction occurs, creating an LF configuration in which Principle C is violated. (11b), however, does not reconstruct; Principle C is not violated. Both Huang (1993) and Heycock (1995) argue that the distinction reduces to a difference in LF position, although they differ on the nature of the distinction. Antireconstruction effects thus provide a diagnostic for whether reconstruction has or has not occurred. Thus, for B&T antireconstruction effects such as the one

<sup>13</sup> Following Comrie (1973), Müller and Sternefeld (1993), and many others since, I assume that extraction in Russian is generally acceptable only out of *čtoby* (subjunctive) clauses, and not out of *čto* (indicative) clauses. Thus, Russian examples of long-distance scrambling will be given in such structures. How the *čtoby* versus *čto* distinction is to be handled is not crucial to the present discussion (but see Avrutin and Babyonyshev 1994 for an interesting account).

<sup>14</sup> An *LJ* reviewer raises the question of a distinct process of (overt) QR for (10b). However, covert QR is required for Russian, where scope ambiguities are common (and overt QR is rare). And because QR is generally taken to be clause-bounded, overt QR could not account for (10b). Therefore, (10b) involves scrambling (lowering for B&T), and it should manifest embedded scope, contrary to speaker judgments.

illustrated in (11b) should disappear with scrambling. However, the same contrast shown in (11) obtains with Russian scrambling.

- (12) a. On [dovol'noj rabotoj Marii]<sub>k</sub> sčitaet ee<sub>j/\*i</sub> t<sub>k</sub> davno.  
 he-TOP satisfied-FEM work-INSTR Mary-GEN considers her long since.  
 'He has considered her<sub>i</sub> satisfied with Mary's<sub>i</sub> work long since.'
- b. On [sluxi o Marii]<sub>k</sub> xočet, čtoby ona<sub>i</sub> uslyšala t<sub>k</sub>.  
 he-TOP rumors about Mary wants that she heard  
 'He wants her<sub>i</sub> to hear rumors about Mary<sub>i</sub>.'
- c. [Nekotorye voprosy Goru]<sub>j</sub> ja xoču, čtoby on<sub>i</sub> srazu zabyt t<sub>j</sub>.  
 some questions Gore-DAT I want that he soon forgot  
 'Some questions to Gore<sub>i</sub> I want him<sub>i</sub> to immediately forget.'

The predicate in (12a) reconstructs, triggering a Principle C violation.<sup>15</sup> The arguments in (12b–c), on the other hand, are interpreted in their surface position; there is no Principle C violation. This is the opposite of what we would expect under B&T's proposal. Given obligatory lowering we should not expect any antireconstruction effects with scrambling (prediction A) since all scrambled arguments *obligatorily* lower for  $\theta$ -checking.<sup>16</sup> However, such effects exist in Russian, as shown in (12). On the other hand, a theory that allows raising to scrambled positions (followed by reconstruction) has the flexibility to handle these data. We now turn to reconstruction in raising accounts.

2.1.3 "Sensitive" Reconstruction Heycock (1995) discusses the reconstruction facts in (11) and concludes that reconstruction is obligatory for "nonreferential" phrases and optional or impossible for others (primarily arguments). This is reminiscent of "optional" movement processes in Germanic (especially object shift), which are related to semantic effects by Diesing and Jelinek (1996) and others, whereby the LF "type" of an optionally shifted element determines its LF position (inside or outside its home VP, for example). If LF reconstruction after movement is "sensitive" to the semantic nature of the element, as such accounts maintain, we can allow for the distinctions found above while maintaining a classical raising account. It is unclear how B&T's account would accommodate semantic sensitivity of this kind since all arguments (and only arguments) must lower at LF to  $\theta$ -position. Clearly, some are not interpreted in  $\theta$ -position.<sup>17</sup>

<sup>15</sup> Notice that the existence of a Principle C violation here is a problem under prediction C as well, since the constituent involved is not an argument and yet appears to reconstruct/lower.

<sup>16</sup> Lebeaux (1988), Heycock (1995), Safir (1999), and many others discuss the possibility of a process of "late insertion" for adjuncts that would allow them to be located high in the structure even in cases where the argument they adjoin to is interpreted in lower position. Late insertion will not be enough, however, to explain away all antireconstruction effects, even if the relevant constituents in (12) are shown to be true adjuncts, something that is not immediately clear. Moreover, Safir shows that there are cases in which *both* the upper and lower copy are required for interpretation, something B&T's account does not appear to predict to be possible.

<sup>17</sup> Here and elsewhere B&T would have to resort to introducing another, different kind of LF level (Huang's LF', for example) to allow both lowering and location in a different position at the time the binding theory applies (a possibility B&T do not discuss). (The same holds for scope, as shown in section 2.1.1.) In effect, this would amount to reintroducing the scrambled position as relevant to the interface, and thus would undermine the spirit of B&T's own analysis.

This accords with a general direction in the literature on LF going back to Chomsky 1976, 1981, and also found in Diesing 1992, Huang 1993, Collins and Thráinsson 1996, and elsewhere. The existence of ‘‘sensitive’’ reconstruction appears to be incompatible with B&T’s LF lowering approach without further elaboration of the accompanying semantics.

## 2.2 *Scrambling as Raising: Syntactic Constraints*

Let us now turn to prediction B—namely, that under B&T’s account no raising is involved in the derivation of alternative word orders.

2.2.1 *Scrambling out of Scrambled Phrases* B&T argue that scrambling out of scrambled phrases is possible because scrambling involves a trivial chain, out of which further movement is acceptable. Recall (7), repeated as (13).

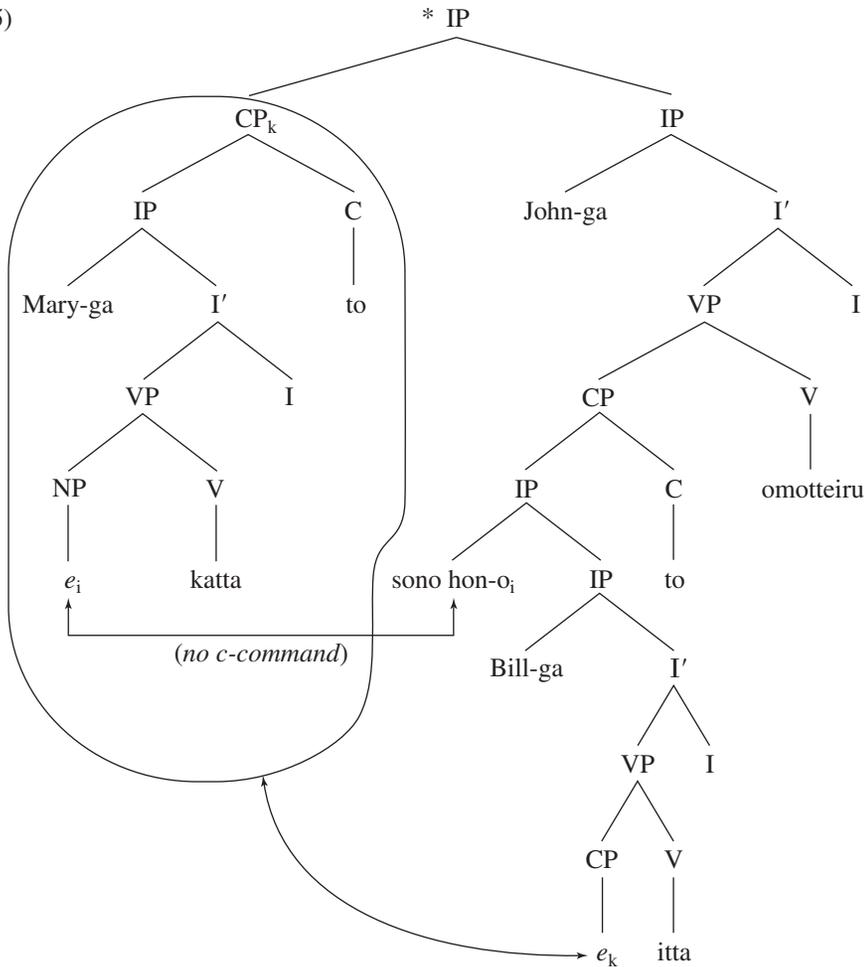
- (13) a. John-ga [Mary-ga *sono hon-o* katta to] omotteiru.  
 John-NOM Mary-NOM that book-ACC bought that thinks  
 ‘John thinks that Mary bought the book.’
- b. [*Sono hono-o<sub>i</sub>* [John-ga [<sub>CP</sub>[<sub>IP</sub>[<sub>CP</sub> Mary-ga  $e_i$  katta to]<sub>k</sub> [Bill-ga  $e_k$  itta]]  
 that book-ACC John-NOM Mary-NOM bought that Bill-NOM said  
 to] omotteiru]].  
 that thinks  
 ‘That book<sub>*i*</sub>, John thinks that [that Mary bought  $e_i$ ]<sub>k</sub>, Bill said  $e_k$ .’

B&T’s account is straightforward: (13b) involves two instances of base generation followed by lowering. Both the smaller (embedded NP) and larger (CP) components are base-generated in adjoined position and undergo LF lowering. Because there are no traces at all, the derivation is licit. ‘‘In [(13b)] . . . no relevant extraction takes place, since both ‘scrambled’ phrases are base-generated where they are at S-Structure’’ (p. 357). Notice, however, that a raising approach fares no worse here. The chains formed by overt movement deriving (13b) violate no derivational principles in a raising account either. In particular,  $e_k$  is c-commanded by its antecedent, as is  $e_i$ . Where the two accounts differ is with respect to a different derivation: one in which the two displaced constituents, CP<sub>k</sub> and NP<sub>i</sub>, appear *in the opposite order*. In this case the lowering analysis predicts grammaticality (in fact, no order should be worse than any other), whereas the raising analysis predicts a violation. The latter prediction is confirmed by the ungrammaticality of (14).

- (14) \*[[<sub>CP</sub> Mary-ga  $e_i$  katta to]<sub>k</sub> [John-ga [<sub>CP</sub> *sono hon-o<sub>i</sub>* [<sub>IP</sub> [Bill-ga  $e_k$  itta]]  
 Mary-NOM bought that John-NOM that book-ACC Bill-NOM said  
 to] omotteiru]].  
 that thinks  
 ‘[That Mary bought  $e_i$ ]<sub>k</sub>, John [that book<sub>*i*</sub>] thinks that, Bill said  $e_k$ .’

Under a raising analysis the ungrammaticality of (14) is the expected result because there is an improper chain between [<sub>NP</sub> that book] and its trace, ruled out by the Proper Binding Condition, as shown in (15).

(15)



Under the lowering analysis the order found in (14) is predicted to be grammatical because the CP and NP lower into argument position, and no Proper Binding Condition violation is possible (on B&T's assumption that this kind of lowering leaves no trace). The sentence should fare no worse than (13b). It is not clear how B&T's proposals can account for the contrast between (13b) and (14).<sup>18</sup>

<sup>18</sup> (14), though ungrammatical, is reminiscent of German remnant movement constructions, which are acceptable. However, the acceptability of such constructions does not weaken the difficulty (14) presents for B&T's account (indeed, their account predicts such cases to be the norm); that is, it predicts any case of scrambling of a larger XP over a more locally scrambled smaller piece from within XP to be acceptable, contrary to fact, and contrary to the prediction made by the Proper Binding Condition. Furthermore, German remnant movement may be restricted to VP remnants, which would not fall under B&T's account in any event, since VPs are not  $\theta$ -marked constituents. See Epstein et al. 1998 for further discussion.

2.2.2 *Wh-Movement and Scrambling Parallels* B&T's lowering account of long-distance scrambling denies the existence of a "record" in base-generated (adjoined) position; in this way LF lowering is allowed. Thus, we do not expect (typical) movement constraints to hold of long-distance scrambling under B&T's analysis (prediction B), assuming either that constraints hold of chains formed between traces and their antecedents or that they apply to the structural position of traces left by movement. And yet there is evidence that such constraints do apply to scrambling, and with considerable regularity, providing additional support for the classical claim that long-distance scrambling is overt  $\bar{A}$ -movement. Consider the Russian sentences in (16).

- (16) a. \**Kogo*<sub>i</sub> Marina znaet [čto [Ivan ljubit *t*<sub>i</sub>]]?  
 who-ACC Marina-NOM knows that Ivan-NOM loves  
 'Who does Marina know that Ivan loves?'  
 b. \**Borisa*<sub>i</sub> Marina znaet [čto [Ivan ljubit *t*<sub>i</sub>]].  
 Boris-ACC Marina-NOM thinks that Ivan-NOM loves  
 'Marina thinks that Ivan loves *Boris*.'

(16a) shows that *wh*-movement is ungrammatical out of an indicative embedded clause. (16b) shows that the same restriction applies to scrambling.<sup>19</sup>

Let us consider possible approaches to such parallels. First, it is possible that (16b) *does not* involve scrambling. If it involves no movement, the (ungrammaticality) parallel with *wh*-movement is unexpected.<sup>20</sup> If it is another kind of movement (not scrambling), as an *LI* reviewer suggests, then the data are being excluded as belonging to another class of phenomena and a significant generalization is lost (either the *wh*-movement/scrambling parallels or a unified account of scrambling). The alternative remains that (16b) *is* a case of scrambling, and therefore under B&T's account it involves lowering. Something about this lowering, therefore, causes the violation. However, B&T claim that the *only* movement in such cases is lowering (which leaves no trace), and in that case parallels such as the one in (16) are unexpected. In Russian (as in German; see Webelhuth 1989) *wh*-movement and scrambling are restricted in highly parallel fashion. A partial list of relevant constraints restricting movement is provided in (17) (see Fowler 1987 and Bailyn 1995 for discussion).<sup>21</sup>

<sup>19</sup> The exact nature of the violation with extraction is the subject of some dispute. See Comrie 1973, Avrutin and Babyonyshev 1994, and Bailyn 1995 for various views. Solving this problem is beyond the scope of this article, but I take the parallel violations in (16) as evidence that upward movement has taken place in both cases.

<sup>20</sup> There are constructions that appear to be grammatical versions of (16b), primarily with nominative arguments. Indeed, B&T cite such an example, (22b), from Müller and Sternefeld (1993), which goes back to Comrie 1973, to show that *wh*-movement and scrambling are not parallel. However, I show in Bailyn 1995 that these are not cases of scrambling, but cases of left-dislocation with a null resumptive pronoun, which explains why nominative morphology is strongly preferred. Without it, as in (16), left-dislocation is unavailable, scrambling remains the only possibility, and the effect of movement constraints remains.

<sup>21</sup> In a footnote B&T set aside the issue of (some of) these constraints by claiming that they are potentially not movement constraints at all: "We ignore here the Coordinate Structure Constraint, the Left Branch Condition, and the Specificity Condition, since it is not at all clear that these are movement constraints" (p. 358, fn. 17). I will not address that possibility further here until the case for the sudden exclusion of these well-known constraints is better made. Since their effects on *wh*-movement are well known, I will continue to treat them in standard fashion.

- (17) *Movement constraints active in Russian*
- Subjacency/Complex NP Constraint
  - Constraint on extraction out of adnominal genitives
  - Coordinate Structure Constraint
  - Empty Category Principle (includes *that-t* effect)
  - Condition on Extraction Domain

Examples of Subjacency violations are shown in (18).

- (18) a. \**Kogo*<sub>i</sub> ty pozvonil agentu, kotoryj ljubit *t*<sub>i</sub>?  
 whom-ACC you-NOM phone spy-DAT who loves  
 ‘Whom did you phone a spy who loves?’
- b. \**Borisa*<sub>i</sub> ty pozvonil agentu, kotoryj ljubit *t*<sub>i</sub>!  
 Boris-ACC you-NOM phone spy-DAT who loves  
 ‘It’s BORIS you phoned a spy who loves!’

(18a) is a Subjacency violation caused by *wh*-movement. (18b) is a Subjacency violation caused by scrambling. If scrambling involves lowering leaving no trace, such a parallel is predicted not to exist. Another example involves the Coordinate Structure Constraint, which holds in strict form in Russian (and probably universally). Like Subjacency in (18), this constraint applies equally to scrambling and to *wh*-extraction.

- (19) a. \**Kogo*<sub>i</sub> ty xočeš’, čtoby Ivan videl [*t*<sub>i</sub> i Mašu]?  
 whom-ACC you want that Ivan saw and Masha-ACC  
 ‘\*Whom do you want Ivan to see and Masha?’
- b. \**Borisa*<sub>i</sub> ty xočeš’, čtoby Ivan videl [*t*<sub>i</sub> i Mašu].  
 Boris-ACC Ivan want that Ivan saw and Masha-ACC  
 ‘\*BORIS you want Ivan to see and Masha.’

In (19a) *wh*-movement violates the Coordinate Structure Constraint. (19b) is equally ungrammatical. This parallel cannot be captured under an obligatory lowering account: in (19b) *Borisa* should be generable in IP-adjoined position and then lower into  $\theta$ -position, without causing any violation, again because no trace is left and no chain is formed. In fact, however, the sentence is just as ungrammatical as the parallel *wh*-extraction case (19a). Examples of this kind abound (see Bailyn 1995 for Russian and Webelhuth 1989 for German). B&T’s analysis appears to leave such parallels unexplained.

### 2.3 *Nonarguments*

B&T’s theory makes another strong testable prediction, prediction C—namely, that only elements assigned a  $\theta$ -role should be able to scramble.<sup>22</sup> If  $\theta$ -role ‘checking’ is the driving force behind the ability to appear displaced from base position, then *only* those elements that receive  $\theta$ -roles

<sup>22</sup> See footnote 10 for discussion of whether features other than  $\theta$ -roles could drive the lowering process. If they could, then B&T’s typology of scrambling breaks down.

should show variation in word order. Non- $\theta$ -marked items should never be found dislocated from their immediate clause. However, Russian demonstrates scrambling of adjectives, adverbs, and other nonarguments.

- (20) a. Ja xoču, čtoby oni bystro dopisali kursovye.  
 I want that they quickly wrote papers  
 ‘I want them to write their papers quickly.’
- b. Ja *bystro* xoču, čtoby oni *t* dopisali kursovye.  
 I quickly want that they wrote papers  
 ‘I want them to write their papers quickly.’
- c. \*Ja *bystro* znaju agenta, kotoryj pišet *t*.  
 I quickly know agent who writes  
 ‘I quickly know an agent who writes.’
- d. Ja *zelenuju<sub>i</sub>* xoču, čtoby ty kupila [*t<sub>i</sub>* knigu].  
 I green-ACC want that you bought book-ACC  
 ‘I want you to buy a green book.’

In (20b) the embedded manner adverb *bystro* ‘quickly’ is located in the main clause, yet it modifies the embedded verb just as in (20a). Under B&T’s theory the acceptability of (20b) is unexpected, without additional assumptions. To be interpreted in the lower clause, *bystro* must be in that position at LF. And yet there is no (feature-driven) motivation for such LF movement. Prediction C is disconfirmed. (20c), a standard Subjacency violation, provides additional evidence for the raising account of scrambling even in cases involving adverbs. (20d) is a case of so-called split scrambling whereby an attributive adjective from the embedded clause is scrambled to the main clause, a movement unexpected under B&T’s account.

### 3 Some Theoretical Issues Raised by Lowering

In addition to empirical issues, the theory proposed by B&T raises theoretical points involving  $\theta$ -theory and feature checking, as well as the original issue of optionality and economy.

#### 3.1 $\theta$ -Theory, Feature Checking, and Multiple Scrambling

Under standard minimalist assumptions (Chomsky 1995),  $\theta$ -marking occurs at Merge, but this requirement is simply stipulated. It appears, however, that any theory of  $\theta$ -marking that does not associate it with Merge greatly increases the empirical burden put on theories of phrase structure, which specifically rule out movement into complement position. If thematic relations are not local at Merge, even those between a head and its complement, the traditional arguments that sisterhood is the closest syntactic relation are lost.

Indeed, in Chomsky 1995 it is made quite clear that  $\theta$ -roles are not formal features: ‘‘Under any approach that takes Attract/Move to be driven by morphological features . . . *there should be no interaction between  $\theta$ -theory and the theory of movement*’’ (p. 312, emphasis mine). Some arguments in favor of this are the following: (a) The Chain Condition defines a chain, formed

by Move, as meeting several conditions “which we take to be part of the definition of the operation [Move] itself. . . .  $\alpha$  must c-command its trace, so that there cannot be an operation that lowers  $\alpha$  . . . ; *movement is raising*, in the specific sense defined by c-command” (p. 253, emphasis mine). The tail of an A-chain is defined by its status as a  $\theta$ -position; under B&T’s theory the notion of an A-chain would then be different for scrambling and nonscrambling languages. (b) Internal arguments are merged into complement position of their predicates. However, complements are crucially *not* within the checking domain of a head. Clearly, in a theory such as B&T’s  $\theta$ -relatedness will (often) not result from Merge at the base at all (in Japanese-type languages, at least) but will be subject to checking in complement position. We might therefore expect arguments to be regularly generated in adjoined positions and to assume their  $\theta$ -positions only at LF, predicting multiple scrambling to be the norm as there should be no preference for arguments to appear in  $\theta$ -position.<sup>23</sup> In fact, multiple long-distance scrambling in Russian is not only rare—it is ungrammatical for many speakers, something the base generation account does not predict.

- (21) a. Ivan        xočet, čtoby Boris peredal *kassetu*        Saše.  
           Ivan-NOM wants that Boris gave cassette-ACC Sasha-DAT  
           ‘Ivan wants Boris to give the cassette to Sasha.’  
       b. Ivan xočet, čtoby *kassetu*<sub>i</sub> Boris peredal *t*<sub>i</sub>        Saše.  
       c. Ivan xočet, čtoby *Saše*<sub>k</sub> Boris peredal *kassetu* *t*<sub>k</sub>.

(21b–c), with one argument scrambled, do not distinguish between the two theories. However, any attempt to scramble both arguments leads to a violation.

- (22) a. \*Ivan *Saše*<sub>k</sub> *kassetu*<sub>i</sub> xočet, čtoby *t*<sub>i</sub> *t*<sub>k</sub> Boris peredal *t*<sub>i</sub> *t*<sub>k</sub>.  
       b. \*Ivan *kassetu*<sub>i</sub> *Saše*<sub>k</sub> xočet, čtoby *t*<sub>i</sub> *t*<sub>k</sub> Boris peredal *t*<sub>i</sub> *t*<sub>k</sub>.

Under the base generation and lowering account, which involves no traces (or chain formation), such constructions should be commonplace, and certainly no less acceptable than (21); whereas under a raising account (22a–b) can be ruled out by Relativized Minimality.<sup>24</sup> Thus, we have both empirical and conceptual reasons not to abandon the idea that  $\theta$ -role assignment is unrelated to movement.<sup>25</sup>

### 3.2. Optionality and Economy

Perhaps the most serious theoretical issue for B&T’s account concerns the very problem they begin with: optionality. Recall the main motivation for B&T’s analysis: to eliminate the apparent

<sup>23</sup> It is possible that there are languages that allow only this kind of dislocation strategy. They would not, however, have the properties of Russian and Japanese.

<sup>24</sup> The proper account of this violation assumes Rizzi’s (1990) Relativized Minimality, under which  $\bar{A}$ -movement over an occupied  $\bar{A}$ -position is a violation. Such restrictions appear hard to capture in B&T’s account.

<sup>25</sup> Inherent Case assignment, which relates lexical Case marking to particular  $\theta$ -roles, also raises serious questions for B&T’s account as it stands. How can this relation be handled if inherently Case-marked elements scramble (which is possible in both Japanese and Russian)? A full account of these difficulties and their potential solution, however, requires a theory of inherent Case marking under minimalism that is beyond the scope of this article.

optionality associated with scrambling. Notice, however, that in solving the Last Resort problem through base generation and obligatory lowering, this analysis creates a new problem involving optionality. Optionality becomes unresolvable in a new way, for although the LF movement itself is motivated, a theory that allows more than one potential Merge position for arguments creates a new problem: optionality at Merge. Under B&T's theory languages with weak  $\theta$ -roles have the *option* of base-generating arguments in either of two positions; optional *movement* may be eliminated, but *optionality* remains, being simply transferred to the base structure. Consider, for example, the numerations associated with (1a) and (1b). They are identical, on standard assumptions. Now consider the operations involved in deriving (1a) and (1b) from this numeration. Clearly, (1b) involves an additional step: the LF lowering, which is absent in (1a). Thus, (1b) should be blocked by (1a) in all cases by (global) economy considerations. B&T therefore adopt Collins's (1997) local economy view that prohibits look-ahead. However, if we are to maintain a strong theory of economy, we should consider the possibility that (1a) and (1b) differ in some way and hence do not have the same numeration, a conclusion that would also solve the initial optionality issue. The numeration underlying (1b) must contain certain information that forces movement. This is in the spirit of recent accounts of long-distance scrambling to which we return in section 5. It is also behind current accounts of object shift (Collins and Thráinsson 1996) as well as accounts of local scrambling (Miyagawa, forthcoming, Bailyn, forthcoming), to which we turn next.<sup>26</sup>

#### 4 Local Scrambling

Thus far we have only examined cases of  $\bar{A}$ -scrambling, which are the core cases covered by B&T. Indeed, B&T themselves focus on long-distance scrambling cases. For local (short) scrambling, B&T's base generation proposal does not involve lowering. Rather, they maintain base generation in IP-adjoined position in cases of local scrambling as well, *with no further movement*. This is consistent with the fact that local scrambling appears not to reconstruct, changes scope relations, and so on. To handle  $\theta$ -checking in these cases, B&T propose that these arguments check  $\theta$ -features in situ at LF, meaning they never occupy  $\theta$ -position. In particular, B&T "hypothesize that when moved to I, a verb may  $\theta$ -mark its object in the IP-adjoined position, allowing it to stay there at LF" (p. 360). This leads B&T into certain problems regarding apparent reconstruction effects with Principle C, which they resolve by redefining the segment created by the initial IP-adjunction.

I will not take a strong stand on this account of local scrambling, primarily because the advantage of claiming base generation in such cases rests on the lowering account of  $\bar{A}$ -scrambling, which we have seen to be problematic. However, it is worth pointing out that many of the difficulties associated with the lowering account of  $\bar{A}$ -scrambling remain for these cases as well.

<sup>26</sup> Collins (1997) considers a similar issue in the derivation of locative inversion constructions in English, concluding that Economy is a local, not global, condition. So long as Last Resort and Minimality are observed, Economy is not violated. However, since B&T's theory for the first time introduces optionality with regard to where arguments are base-generated in Japanese, the optionality problem (choosing IP-adjoined position over  $\theta$ -position) remains.

First, there is the issue of rigidity of interpretation, in this case associated with the higher position, since the argument is claimed never to be in  $\theta$ -position. Indeed, Brown (2000) gives evidence from Russian for reconstruction of A-chains, which provides a counterexample to B&T's account of local scrambling. Again, "sensitive" reconstruction appears to be empirically closer to the mark. Second, it is predicted again that (local) scrambling should also be limited to  $\theta$ -marked elements, a prediction that appears to be too strong for Russian, where adverbials, adjectives, and other nonarguments can scramble locally, as shown in (23).

- (23) a. *Bystro*<sub>i</sub> Ivan *t*<sub>i</sub> čitaet knigu.  
 quickly Ivan reads book  
 'Quickly Ivan reads the book.'
- b. [*O Nabokove*]<sub>i</sub> Ivan včera kupil knigu *t*<sub>i</sub>.  
 about Nabokov Ivan yesterday bought book  
 'About Nabokov Ivan bought a book yesterday.'
- c. Ivan *xorošuju*<sub>i</sub> kupil [*t*<sub>i</sub> knigu].  
 Ivan good-ACC bought book-ACC  
 'Ivan bought a good book.'

Third, there is evidence from both Japanese and Russian that local scrambling is not distinct from movement satisfying the EPP. Thus, both Miyagawa (forthcoming) and I (Bailyn forthcoming) analyze local argument scrambling as EPP-driven movement applying to constituents other than the canonical subject. Fourth, it is unclear how Case checking can work, if locally scrambled arguments do not lower even at LF. How, in that case, would a locally scrambled, say, direct object be associated with the appropriate position where its Case is checked? Fifth, the same scope and binding difficulties discussed above will apply to local scrambling cases as well. Sixth, under a base generation account there again appears to be no reason not to left-adjoin *all* internal arguments rather than (sometimes) base-generating them in actual  $\theta$ -position, since the mechanism for  $\theta$ -role assignment to the higher site is available. Thus, we can see that the base generation analysis for local scrambling will also require further work before it can be generally accepted.

## 5 An Alternative Account of Apparent Optionality

### 5.1 Possible Approaches

Let us return again to the optionality involved in (1a) and (1b). Logical possibilities for analyzing such apparent optionality are given in (24).

- (24) *Possible accounts of (1a) and (1b)*
- a. (1a) and (1b) are not derivationally related.
- i. Some languages allow nonconfigurational structures. (nonconfigurationality: Hale 1983)
- ii. Obligatory lowering accounts for (1b). (B&T 1998)
- b. (1a) is derived from (1b). (requires overt lowering)

- c. (1b) is derived from (1a).
  - i. The raising is truly optional, for independent reasons. (Saito and Fukui 1998)
  - ii. The raising is not optional; instead, it is related to another part of the grammar. (Zubizarreta 1998; this article)

(24ai) is assumed to be false in much work on languages like Russian and Japanese and is also incompatible with constraints on scrambling. (24a<sub>ii</sub>) is B&T's approach. (24b) would amount to a B&T-style account of English as well, under which an object in canonical object position is overtly lowered (forced by strong  $\theta$ -features). However, this would require loosening constraints on overt lowering, with obvious empirical difficulties. (24ci) is Saito and Fukui's (1998) account whereby some movement is "truly optional," in the sense that it is not driven by feature checking, but is costless depending on the directionality of phrase structure, so that a left-headed language allows optional movement to the right, whereas only right-headed languages allow left-(IP)-adjunction scrambling. However, this account is clearly too narrow to accommodate right-branching languages with left-adjunction scrambling, such as Russian or Serbo-Croatian. Stjepanović (1998) demonstrates convincingly that this approach will simply not allow enough variation to account for such languages. Only (24cii), a (nonoptional) raising account, remains. Such an approach would account for the parallel behavior noted above. It would also solve many of the conceptual problems discussed earlier. The central question becomes, then, what motivates the movement that raises a constituent out of its position at Merge to a distinct surface location? Such movement does not take place for Case reasons, or for  $\theta$ -reasons. There is no relevant operator-variable chain, and therefore the typical feature-driven accounts will not hold. It is to this issue that I now turn.

## 5.2 Behind the Optionality of Scrambling

In this section I indicate a possible direction for solving the paradox of apparent optionality of scrambling while still allowing a strong derivational approach to alternative word orders. Following Zubizarreta (1998), I assume that scrambling satisfies Last Resort in being forced by a mismatch between the discourse structure (her Assertion Structure) of the utterance and its intonational structure. Thus, for Zubizarreta, Romance subject postposing is a case of "P(rosodic)-movement" without which the eventual Assertion Structure would be underivable by the given word order and intonation. This is also behind Miyagawa's (1997) proposal that  $\bar{A}$ -scrambling "is motivated by something like focus" (p. 21). I claim that  $\bar{A}$ -scrambling is related to the topic-focus or information structure component of the grammar (see Chomsky 1971, Jackendoff 1973, Rochemont 1980, Vallduví 1992, Lambrecht 1994, Zubizarreta 1998, among many others, for similar ideas).

It is well known that word order variants in Russian are associated with distinct discourse interpretations. The usual description, given in Švedova 1980, says that "word order can vary, but at the same time *it is not free*: the meaning of a sentence, its *communicative goal*, differs with different word orders" (p. 191, emphasis mine). The standard account goes as follows: Varying the word order affects the "functional perspective" of the sentence, with a systematic

relationship holding between rheme and theme. A typical contrast is the one shown in (25)–(26) for Russian.

- (25) a. Kto čitaet knigu?  
 who reads book-ACC  
 ‘Who is reading the book?’  
 b. Knigu čitaet Ivan.  
 book-ACC reads Ivan-NOM  
 ‘IVAN is reading the book.’
- (26) a. Čto delaet Ivan?  
 what does Ivan-NOM  
 ‘What is Ivan doing?’  
 b. Ivan čitaet knigu.  
 Ivan reads book  
 ‘Ivan is reading A BOOK.’

(25a) and (26a) are questions that determine appropriate discourse relations in the answer (see Kovtunova 1976 for discussion of this question test). In particular, (25a) asks about the reader of the book. In the answer, therefore, *Ivan* takes final position (usually allotted to the rheme). In (26), however, the situation is reversed and *the book* is the rheme, hence its position at the end of the sentence. Returning to (1a–b), repeated here, we can see that similar factors are at play in Japanese as well.

- (1) a. John-ga [Mary-ga *sono hon-o* katta to] omotteiru.  
 John-NOM Mary-NOM that book-ACC bought that thinks  
 ‘John thinks that Mary bought the book.’  
 b. *Sono hon-o* John-ga [Mary-ga *t* katta to] omotteiru.

(1b) and (1a) differ in discourse structure in the expected way. In particular, (1b) (but not (1a)) is appropriate in contexts where *the book* is part of the preceding discourse or in the “shared common concern” in the sense of Yokoyama 1986. This is shown in (27)–(28).

- (27) a. John-wa dou shiteiru no?  
 John-TOP how doing Q  
 ‘How is John doing?’  
 b. John-ga [Mary-ga *sono hon-o* katta to] omotteiru.  
 John-NOM Mary-NOM that book-ACC bought that thinks  
 ‘John thinks that Mary bought the book.’  
 c. #*Sono hon-o* John-ga [Mary-ga *t* katta to] omotteiru.
- (28) a. Sono hon ni-kanshite nani-ka atta no?  
 that book about something happened Q  
 ‘Did anything happen to that book?’

- b. #John-ga [Mary-ga *sono hon-o* katta to] omotteiru.  
 John-NOM Mary-NOM that book-ACC bought that thinks  
 ‘John thinks that Mary bought the book.’
- c. *Sono hon-o* John-ga [Mary-ga *t* katta to] omotteiru.

(27) shows that a question introducing *John* as the theme of the discourse elicits (27b) as a response, and not (27c). Conversely, a question such as (28a) that introduces *the book* as thematic elicits (28c) as a response, and not (28b). Thus, the scrambled and nonscrambled orders differ regarding the discourse status of the utterance.<sup>27</sup> Minimalist assumptions force us to take such differences seriously, since they bear directly on the linguistic encoding of the sentence, which only reflects factors relevant to the interfaces. The following generalization regarding long-distance scrambling is a strong claim that I hope will provoke further discussion:

(29) *The scrambling generalization*

- a.  $\bar{A}$ -scrambled and nonscrambled orders are *always* associated with different discourse/informational interpretations.<sup>28</sup>
- b. The movement deriving scrambled orders is *motivated* by discourse/informational considerations.<sup>29</sup>

(29a) is in keeping with traditional discourse grammars. (29b), providing a motivation for long-distance scrambling, is in full keeping with the spirit of B&T’s account. Thus, under (29) long-distance scrambling is not optional. However, (29) associates its obligatoriness with discourse notions, not with  $\theta$ -relations. (29) captures the discourse effects mentioned above, and maintains a raising account.<sup>30</sup> It also avoids the theoretical difficulties discussed in section 3.

Preliminary evidence for (29a) involves the availability of so-called ‘‘functional ambiguity,’’ whereby more than one discourse structure is available given a certain word order. As shown in Lavine 1999, in Russian only canonical SVO orders (or orders derived with A-movement alone) allow functional ambiguity. Thus, the discourse structure of an SVO clause such as (30a) could be any of the structures given in (30b–d).

<sup>27</sup> Of course this distinction is nothing close to a new discovery in linguistics. It has a long tradition in Slavic linguistics, for example, going back at least to Mathesius 1939, appearing throughout Soviet and Prague School linguistics, and being picked up in various forms in many branches of pregenerative linguistics and functional approaches to syntax (see, among many others, Adamec 1966, Babby 1980, Prince 1984, Yokoyama 1986, Kuno 1987, Hajičová and Sgall 1987). Its integration with current understanding of derivational violations appears therefore to represent a step forward in our understanding of human language.

<sup>28</sup> (29a) is limited to  $\bar{A}$ -scrambling because A-scrambling appears to be less discourse-related (if it is discourse-related at all). On the other hand, A-scrambling presents less of an optionality problem because of its possible association with formal features (such as the EPP). See Miyagawa, forthcoming, and Bailyn, forthcoming, for discussion.

<sup>29</sup> The question arises here whether or not discourse-related movement is driven by formal features of Topic and Focus, or whether other factors are at play. This issue is beyond the scope of this article; but for the optionality problem to be solved, the numerations underlying (1a) and (1b) must be distinct, presumably differing by discourse-related information in one form or another. This issue is also left open in Zubizarreta 1998.

<sup>30</sup> Of course, some discourse effects, such as those resulting from left-dislocation, are indeed base-generated (see footnote 20). Others clearly are not. The claim here is unidirectional—namely, that  $\bar{A}$ -scrambling entails a discourse structure distinct from nonscrambled orders, but not that all such distinct discourse structures are derived by  $\bar{A}$ -scrambling.

- (30) a. *čto Ivan čitaet knigu*  
 that Ivan reads book  
 ‘that Ivan reads/is reading a book’  
 b. *Ivan čitaet [knigu]* (subject + verb = topic; NP object = comment)  
 c. *Ivan [čitaet knigu]* (subject = topic; VP = comment)  
 d. *[Ivan čitaet knigu]* (no topic; all comment)

However, once long-distance scrambling has occurred, discourse structure is no longer ambiguous.<sup>31</sup> This is shown in (31).

- (31) a. *čto knigu<sub>i</sub> [Ivan čitaet t<sub>i</sub>]*  
 that book Ivan reads  
 ‘that the book, Ivan reads’  
 b. *[knigu] Ivan čitaet* (scrambled object = topic; subject + V = comment)  
 c. *[knigu Ivan] čitaet* ???(scrambled object + subject = topic; verb = comment)  
 d. *[knigu Ivan čitaet]* \*(no topic; all comment)

How might a theory such as B&T’s represent discourse relations? Given that the scrambled element is associated with the higher position only before LF lowering, discourse relations would have to be represented at the level of the initial phrase marker—that is, at Merge (if such a level could be isolated, contra general assumptions of minimalism). Relations established at this level would be ‘‘undone’’ by obligatory lowering and thus would not be represented at any unified interface level. This difficulty serves as an additional theoretical argument against B&T’s approach, on the assumption that Zubizarreta (1998), Heycock and Kroch (1999), and others are right in associating discourse relations with the interpretive component of the grammar.

## 6 Conclusion

We have seen that B&T’s claim that scrambling involves base generation followed by obligatory LF lowering creates more problems than it solves, empirically and theoretically. Their claim is both too weak and too strong. It is too weak in that it does not allow for constraints on scrambling and does not account for the parallel behavior of *wh*-movement and scrambling, whereas a raising account does. On the other hand, it is too strong in terms of the nature of the LF process involved, which, if morphologically driven, cannot be sensitive to semantic considerations. And yet scrambling in fact seems to pattern with *wh*-movement (and topicalization) in undergoing (semantically sensitive) reconstruction, an unexpected result for B&T’s approach. Their claim also leads us to expect more scrambling than is actually observed and introduces new issues of optionality and economy. Base generation also does not allow for uniform representation of discourse relations at an interface level.

<sup>31</sup> Intonation also plays an important role in determining discourse status in spoken language. However, the relation between intonation and discourse structure is a complicated issue that falls mostly outside the scope of this article. For this example, I assume neutral intonation that follows some version of the Nuclear Stress Rule. For discussion, see Bryzgunova 1971, 1981, Rochemont 1980, Selkirk 1984, and Zubizarreta 1998.

## References

- Adamec, Přemysl. 1966. *Porjadok slov v sovremennom ruskom jazyke*. Prague: Academia.
- Avrutin, Sergei, and Maria Babyonyshev. 1994. Obviation in subjunctive clauses and AGR: Evidence from Russian. In *Formal Approaches to Slavic Linguistics: The Ann Arbor Meeting 1992*, 1–24. Michigan Slavic Publications, University of Michigan, Ann Arbor.
- Babby, Leonard H. 1980. *Existential sentences and negation in Russian*. Ann Arbor, Mich.: Karoma Publishers.
- Bailyn, John Frederick. 1995. A configurational approach to Russian “free” word order. Doctoral dissertation, Cornell University, Ithaca, N.Y.
- Bailyn, John Frederick. Forthcoming. Generalized inversion. *Natural Language & Linguistic Theory*.
- Bayer, Josef, and Jaklin Kornfilt. 1994. Against scrambling as an instance of Move-alpha. In *Studies on scrambling: Movement and non-movement approaches to free word-order phenomena*, ed. Norbert Corver and Henk van Riemsdijk, 17–60. Berlin: Mouton de Gruyter.
- Boeckx, Cedric, and Sandra Stjepanović. 2000. The *wh*/clitic-connection in Slavic: Theoretical implications. In *Formal Approaches to Slavic Linguistics: The Philadelphia Meeting 1999*, 22–40. Michigan Slavic Publications, University of Michigan, Ann Arbor.
- Bošković, Željko. 1997. Fronting *wh*-phrases in Serbo-Croatian. In *Formal Approaches to Slavic Linguistics: The Indiana Meeting 1996*, 86–107. Michigan Slavic Publications, University of Michigan, Ann Arbor.
- Bošković, Željko. 1998. *Wh*-phrases and *wh*-movement in Slavic. Position paper presented at the Comparative Slavic Morphosyntax Workshop, Bloomington, Ind. Available at: <http://www.indiana.edu/~slavconf/linguistics/download.html>.
- Bošković, Željko, and Daiko Takahashi. 1998. Scrambling and Last Resort. *Linguistic Inquiry* 29:347–366.
- Brown, Sue. 2000. Evidence for A-chain reconstruction. Poster presented at the 9th Annual Workshop on Formal Approaches to Slavic Linguistics, Indiana University, Bloomington, February 2000.
- Bryzgunova, Elena. 1971. O smyslorazličitel’nyx vozmožnostjax ruskoj intonacii. *Voprosy jazykoznanija* 4:42–51.
- Bryzgunova, Elena. 1981. *Zvuki i intonacija ruskoj reči*. Moscow: Russkij Jazyk.
- Chomsky, Noam. 1971. Deep Structure, Surface Structure, and semantic interpretation. In *Semantics*, ed. Danny Steinberg and Leon Jakobovits, 183–216. Cambridge: Cambridge University Press.
- Chomsky, Noam. 1976. Conditions on rules of grammar. *Linguistic Analysis* 2:303–351.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, Mass.: MIT Press.
- Collins, Chris. 1997. *Local economy*. Cambridge, Mass.: MIT Press.
- Collins, Chris, and Höskuldur Thráinsson. 1996. VP-internal structure and object shift in Icelandic. *Linguistic Inquiry* 27:391–444.
- Comrie, Bernard. 1973. Clause structure and movement constraints in Russian. In *You take the high node and I’ll take the low node: Papers from the Ninth Regional Meeting, Chicago Linguistic Society*, 291–304. Chicago Linguistic Society, University of Chicago, Chicago, Ill.
- Diesing, Molly. 1992. *Indefinites*. Cambridge, Mass.: MIT Press.
- Diesing, Molly, and Eloise Jelinek. 1996. Distributing arguments. *Natural Language Semantics* 2:1–54.
- Epstein, Samuel David, Erich M. Groat, Ruriko Kawashima, and Hisatsugu Kitahara. 1998. *A derivational approach to syntactic relations*. New York: Oxford University Press.
- Erteschik-Shir, Nomi. 1997. *The dynamics of focus structure*. Cambridge: Cambridge University Press.
- Fowler, George. 1987. The grammatical relevance of theme/rheme partition. In *CLS 23*, 93–104. Chicago Linguistic Society, University of Chicago, Chicago, Ill.
- Hajičová, Eva, and Petr Sgall. 1987. The Ordering Principle. *Journal of Pragmatics* 11:435–454.

- Hale, Ken. 1983. Warlpiri and the grammar of non-configurational languages. *Natural Language & Linguistic Theory* 1:5–47.
- Heycock, Caroline. 1995. Asymmetries in reconstruction. *Linguistic Inquiry* 26:547–570.
- Heycock, Caroline, and Anthony Kroch. 1999. Pseudocleft connectedness: Implications for the LF interface level. *Linguistic Inquiry* 30:365–397.
- Huang, C.-T. James. 1993. Reconstruction and the VP-Internal Subject Hypothesis. *Linguistic Inquiry* 24:103–138.
- Jackendoff, Ray. 1973. *Semantic interpretation in generative grammar*. Cambridge, Mass.: MIT Press.
- Jungmanns, Uwe, and Gerhilde Zibatow. 1995. Syntax and information structure of Russian clauses. Paper presented at Formal Approaches to Slavic Linguistics 4, Cornell University, Ithaca, N.Y.
- Kovtunova, Irina. 1976. *Porjadok slov i aktual'noe členenie predloženiya*. Moscow: Prosveščenie.
- Kuno, Susumu. 1987. *Functional syntax: Anaphora, discourse, and empathy*. Chicago: University of Chicago Press.
- Lambrecht, Knut. 1994. *Information structure and sentence form*. Cambridge: Cambridge University Press.
- Lavine, James. 1999. Subject properties and ergativity in North Russian and Lithuanian. In *Formal Approaches to Slavic Linguistics: The Seattle Meeting 1998*, 307–328. Michigan Slavic Publications, University of Michigan, Ann Arbor.
- Lebeaux, David. 1988. Language acquisition and the form of grammar. Doctoral dissertation, University of Massachusetts, Amherst.
- Lee, Y.-S., and Beatrice Santorini. 1994. Towards resolving Weibelhuth's paradox: Evidence from German and Korean. In *Studies on scrambling: Movement and non-movement approaches to free word-order phenomena*, ed. Norbert Corver and Henk van Riemsdijk, 257–300. Berlin: Mouton de Gruyter.
- Mahajan, Anoop. 1990. The A/A' distinction and movement theory. Doctoral dissertation, MIT, Cambridge, Mass.
- Mathesius, Vilém. 1939. O tak zvaném aktualním členění věty. In *Slovo a slovesnost: Ročník V*, ed. Vilém Mathesius and Bohuslav Havránek, 171–174. Prague: Melantricha A.S.
- Miyagawa, Shigeru. 1997. Against optional scrambling. *Linguistic Inquiry* 28:1–26.
- Miyagawa, Shigeru. Forthcoming. EPP, scrambling, and *wh* in situ. In *Scrambling and word order*, ed. Simin Karimi. Oxford: Blackwell.
- Müller, Gereon, and Wolfgang Sternefeld. 1993. Improper movement and unambiguous binding. *Linguistic Inquiry* 24:461–509.
- Neeleman, Ad. 1994. Scrambling as a D-Structure phenomenon. In *Studies on scrambling: Movement and non-movement approaches to free word-order phenomena*, ed. Norbert Corver and Henk van Riemsdijk, 387–430. Berlin: Mouton de Gruyter.
- Prince, Ellen. 1984. Topicalization and left-dislocation: A functional analysis. In *Discourses in reading and linguistics*, ed. Sheila J. White and Virginia Teller, 213–225. Annals of the NY Academy of Sciences 443. New York: NY Academy of Sciences.
- Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, Mass.: MIT Press.
- Rochemont, Michael. 1980. *Focus in generative grammar*. Philadelphia: John Benjamins.
- Ross, John Robert. 1967. Constraints on variables in syntax. Doctoral dissertation, MIT, Cambridge, Mass.
- Safir, Ken. 1999. Vehicle change and reconstruction in  $\bar{A}$ -chains. *Linguistic Inquiry* 30:587–620.
- Saito, Mamoru. 1989. Scrambling as semantically vacuous  $\bar{A}$ -movement. In *Alternative conceptions of phrase structure*, ed. Mark Baltin and Anthony Kroch, 182–200. Chicago: University of Chicago Press.
- Saito, Mamoru. 1992. Long distance scrambling in Japanese. *Journal of East Asian Linguistics* 1:69–119.
- Saito, Mamoru, and Naoki Fukui. 1998. Order in phrase structure and movement. *Linguistic Inquiry* 29:439–474.
- Selkirk, Elisabeth. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, Mass.: MIT Press.

- Stepanov, Artur. 1998. On *wh*-fronting in Russian. In *NELS 28*, 453–467. GLSA, University of Massachusetts, Amherst.
- Stjepanović, Sandra. 1998. *Scrambling in Serbo-Croatian*. Ms., University of Connecticut, Storrs.
- Strakhov, Natalia. 2001. A scrambling analysis of *wh*-questions in Russian. In *Formal Approaches to Slavic Linguistics: The Bloomington Meeting 2000*. Michigan Slavic Publications, University of Michigan, Ann Arbor.
- Švedova, Natalia, ed. 1980. *Russkaja grammatika*. Vol. 2. Moscow: Nauka.
- Vallduví, Enric. 1992. *The informational component*. New York: Garland.
- Webelhuth, Gert. 1989. Syntactic saturation phenomena and the modern Germanic languages. Doctoral dissertation, University of Massachusetts, Amherst.
- Yadroff, Mikhail. 1992. The syntactic properties of adjunction (scrambling in Russian). Ms., Indiana University, Bloomington.
- Yokoyama, Olga. 1986. *Discourse and word order*. Philadelphia: John Benjamins.
- Zubizarreta, Maria Luisa. 1998. *Prosody, focus, and word order*. Cambridge, Mass.: MIT Press.

Department of Linguistics  
 Social and Behavioral Sciences  
 SUNY at Stony Brook  
 Stony Brook, New York 11794-4376  
 jbailyn@notes.cc.sunysb.edu

## Zero Derivations

*Luigi Burzio*

Rubach (2000) proposes a modified version of Optimality Theory (OT) that features derivations. While Prince and Smolensky's (1993) original formulation requires some modification, I argue here that, rather than reintroducing derivations, the correct approach is to take fuller advantage of OT's inherent parallelism. I propose that outputs must be related not only to inputs, but to other, "neighboring" representations as well—a feature that is shared by both the output-to-output faithfulness approach and the theory of targeted constraints developed by Wilson (2000, to appear). I show that all the cases cited by Rubach that seem to support derivations are in fact handled by the latter two related theories, and that both of these have significant advantages over derivations.

*Keywords:* derivations, opacity, output-to-output faithfulness, targeted constraints

## 1 Introduction

Rubach (2000) proposes a version of Optimality Theory (OT) that features derivational levels or stages: Derivational OT (DOT), a development also advocated by Kiparsky (1998). In this article

I am grateful to Bob Ladd, Colin Wilson, and two anonymous reviewers for insightful comments and suggestions.

I defend an alternative conception, which has no serial derivations and is instead fully parallel. I have argued elsewhere (Burzio 2000b) that a single modification of Prince and Smolensky's (1993) original formulation provides a unitary solution to a number of outstanding problems, including the ones examined by Rubach. The modification consists of allowing the calculation of an output in OT not just to rely on a unique "input," but to make comparisons with other representations as well, as in the output-to-output faithfulness approach that Rubach argues against. Unlike reintroducing derivations, this modification is a natural one, since it merely exploits the inherent character of OT, namely, its parallelism.<sup>1</sup> To put it differently, it would actually be an accident for OT if outputs were always calculated from a single input, since—in that architecture—they can be calculated from multiple representations simultaneously just as well, by concurrent application of multiple faithfulness constraints. The view that there is a single input is arguably just a carryover from rule-based systems, in which one-to-one mappings were a contingency of the rewrite rules. If one makes a commitment to that element of rule systems, it is not surprising that other elements may be found necessary as well, just as Rubach claims. My aim in this article is to remove the arbitrary premise that mappings are only one-to-one and, with it, the conclusion it entails—that there are derivations.

I will examine the cases that Rubach adduces in support of derivations, dividing them into three categories: (a) cases that can alternatively be handled by means of output-to-output faithfulness (OO-FAITH), as Rubach acknowledges; (b) cases of opacity (counterbleeding/counterfeeding) that can alternatively be handled by some appropriate parallel theory of opacity—a possibility that Rubach also acknowledges by considering McCarthy's (1999) sympathy theory; (c) other cases. I will argue that the last category includes only spurious uses of derivations, that is, cases that are adequately handled even by standard OT means. If this classification is correct, Rubach's argument for derivations does not go beyond the claim that parallel solutions are heterogeneous (output-to-output faithfulness and some other theory (perhaps sympathy) to deal with opacity), while the serial solution is unitary (derivations). I will show that this claim is incorrect for two reasons: (a) there exists a well-motivated theory of opacity, developed by Wilson (2000, to appear), which can be naturally related to output-to-output faithfulness; (b) there is in fact no relationship between the derivational surrogates that Rubach proposes for the above two theories of output-to-output faithfulness and sympathy, other than the relatively trivial one that they are both derivational.

## 2 Output-to-Output Faithfulness Cases

Rubach (p. 289) proposes the following type of derivation for Bulgarian, to be utilized also for Czech and rural Polish:

(1)		<i>Gloss</i>	<i>Critical ranking</i>
	a. Word level	/amerika/ → ʔamerika	'America'      ONSET >> *ʔ
	b. Phrase level	s ʔamerika → s ʔamerika	'with America'      MAX >> *ʔ

<sup>1</sup> Obviously, Rubach's proposed partial serialism is not *logically* impossible within OT. Rather, it only seems conceptually unnatural.

*S* *ʔamerika* in (1b) is a case of opacity: specifically, counterbleeding, since the presence of the preposition *s* ‘with’, which would provide the needed onset, fails to bleed *ʔ*-epenthesis. Although not all cases of opacity can be accounted for by output-to-output faithfulness, some—like this one—can, as Rubach admits. His account is in fact virtually identical to an output-to-output faithfulness account, and no derivation is entailed by it. That is, the two calculations in (1) can be seen as proceeding in parallel without any difficulty. One need only postulate one principled difference between the grammars that apply in the two cases: calculation of phrases is constrained by faithfulness to the form that words have in isolation (the MAX of (1b) to be understood on this view as MAX<sub>OO</sub>), while calculation of words in isolation is generally insensitive to the structure of the phrases that may contain them. This is Benua’s (1997) Base Priority Principle, which relates larger units to their parts, in a nonderivational way.<sup>2</sup> While Rubach’s derivational analysis and the nonderivational output-to-output faithfulness analysis of (1) may just seem inconsequential restatements of one another, other considerations distinguish the two approaches. Consider that phenomena like the one in (1b) are found as well with English level 2 affixes. For example, unlike in level 1 *metr-ic*, which is regular (Borowsky 1993), in level 2 *meteR-ing* the process that causes word-final *r* to be syllabic ‘overapplies’ just as *ʔ*-epenthesis does in (1b). Here, Rubach’s derivational approach would postulate a level 1 stage, feeding level 2. At level 2—and only at level 2—faithfulness will dominate markedness as it does in (1b), ensuring that *meteR-ing* will mirror *meteR*, which is the output of level 1. This is insufficient, however. Level 1 and level 2 formations differ simultaneously along two independent dimensions: phonological (regular application vs. overapplication of phonological processes) and morphological. Unlike level 2 formations, level 1 words exhibit widespread morphological irregularity; this is illustrated in (2), where the capitalized portions are absent in the respective bases, (2e) and (2f) instantiating the phenomenon of bound stems.

(2) *Irregular level 1 morphology (augmented, distorted, and bound stems)*

- |                 |   |              |
|-----------------|---|--------------|
| a. crime        | → | crimIN-al    |
| b. problem + ic | → | problemAT-ic |
| c. horizon + al | → | horizonT-al  |
| d. compel + ive | → | compUIS-ive  |
| e. ??           | → | ARBORE-al    |
| f. ??           | → | INEVIT-able  |

Since such irregularities are systematically absent with level 2 formations (cf. *crime-less*, *compelled*, etc.), the generalization is thus that their stems are strongly required to equal words (Burzio 1994b:274). This requirement has the effect of suppressing both the kind of distortion illustrated for level 1 in (2) and certain phonologically driven alternations like *meteR/metr-ic* or the other

<sup>2</sup> Rubach (fn. 31) finds it ‘uncontroversial’ that words are derivationally prior to phrases. However, it is only the weaker conclusion that words are *epistemologically* prior to phrases that is—relatively—uncontroversial. The stronger conclusion that such priority is expressed by discrete derivational stages would need justification. Even the weaker conclusion can only be relatively uncontroversial, since it seems only partially true. For instance, the first person singular present of the verb *be* is expressed by different words depending on its phrasal context: *Am I not?* versus *Aren’t I?*

familiar ones in (3), all absent at level 2 (witness *cri:me-less* (no shortening), *frolick-ing* (no velar softening), *párent-hood* (no restressing)).

(3) *Regular level 1 phonology*

- a. *Vowel shortening*: *cri:me/crimin-al*, *na:ture/natur-al*, *admi:re/admir-able*, . . .
- b. *Velar softening*: *authentic/authenti[s]-ity*, *critic/criti[s]-ism*, . . .
- c. *Restressing*: *párent/parént-al*, *títan/titán-ic*, . . .

This inverse correlation of phonological and morphological regularities (*irregular* morphology/*regular* phonology for the level 1 lexicon vs. *regular* morphology/*irregular* phonology for level 2), is what the output-to-output faithfulness approach expresses: high-ranked OO-FAITH for the level 2 lexicon will both inhibit regular phonological alternations and suppress morphological idiosyncrasy, while low-ranked OO-FAITH for level 1 will permit both. As for the source of morphological idiosyncrasy, this is taken to be input-to-output faithfulness (IO-FAITH). In the approach developed in Burzio 1994b, et seq., morphologically complex forms are related to their bases only via their surface forms (OO-FAITH), rather than by a shared underlying representation (IO-FAITH). This allows IO-FAITH to be deployed with morphologically derived forms in the same capacity as with morphologically simple ones: to express what is item-specific and unpredictable (see Burzio 2000a, to appear a,b, for details, and Benua 1997 for a partially similar argument). In contrast, in the derivational approach the morphological subpart of the generalization requires the assumption that level 1 may only output actual words (no bound stems, no augmented or distorted words like *crimIN*, etc.). However, that is quite unrelated to the assumption about level 2 phonology needed to express the other half of the generalization. The two separate provisions thus form a conspiracy to the effect that stems of level 2 formations must equal words. Such conspiracies are a textbook argument against derivations and for surface constraints (Kager 1999: 56), in this case supporting output-to-output faithfulness.

In fact, derivations appear irrelevant even to the strictly phonological side of the generalization. What yields the difference between *metr-ic* and *meteR-ing*, or between shortening *crim-in-al* and nonshortening *cri:me-less*, is the assumption that FAITH (whether IO- or OO-) is higher-ranked with level 2 than with level 1 affixes. The sequential derivation itself contributes nothing, just as in the case of the Bulgarian example in (1b).

In sum, derivations are both unnecessary and insufficient: phonological generalizations hinge on high-ranked versus low-ranked FAITH (derivations are unnecessary); morphological generalizations hinge on the notion that faithfulness holds between words (outputs). This is true by definition under a theory that includes OO-FAITH, but needs stipulating under one based on derivations (derivations are insufficient).

Failure to distinguish IO- from OO-FAITH as in Rubach's system has further puzzling consequences, at least in English. The reason is that it is not a generalization about IO-FAITH that its rank increases at higher levels of composition like level 2 and phrases. While FAITH must be lower-ranked for level 1 than for level 2 so as to permit vowel shortening in (e.g.) *crim-in-al*, while blocking it in *cri:me-less*, it must nonetheless be relatively high-ranked for underived items in general so as to block the vowel shortening in those items (e.g., *cri:me*, *vi:tamin*, *di:nosaur*).

If all FAITH is IO, the question arises why its rank varies nonmonotonically over what would otherwise seem a descriptively well motivated progression of morphological domains—underived, level 1, level 2, phrases—that rank going from high, to low, to high again. This puzzle disappears if one recognizes *both* IO- and OO-FAITH. On this view OO-FAITH is simply irrelevant to underived items, which are thus calculated from IO-FAITH alone. In contrast, at least in the conception developed in Burzio 1996, 2000a, to appear a,b, in which there is no underlying representation, derived items are calculated solely from OO-FAITH (aside from the irregularities of (2), which require partial inputs as noted). On this general conception, OO-FAITH can then be seen as re-ranking upward monotonically from level 1 to level 2 to phrases. On this view the problem of nonhomogeneity of the grammar over different morphological domains, which is theory-independent (after all, phonology just works differently over level 1/level 2/phrases) will reduce to the question of why the rank of OO-FAITH rises (now monotonically) as the size of the units grows (phrases being larger than words, and level 2 formations apparently being quasi-phrasal from this perspective). As for the rank of IO-FAITH, it never varies at all. It is simply ranked high enough to block vowel shortening and other processes (Burzio 2000a). The next few remarks address the question of the apparent dependency between unit size and rank of OO-FAITH.

I have argued in Burzio, to appear a,b, that the effects encapsulated in the notion of OO-FAITH are the combination (by some appropriate operation—e.g., summation of their scalar values, or the ‘local conjunction’ of more familiar OT work) of two different effects. The first, which I refer to as *SELECT*, is simply a specific expression of the needed combinatorial principle. The second component, which I term *gradient attraction*, pertains to independent similarity of representations in multidimensional space. Beginning with the first component, consider for instance that *-less* attaches only to nouns. This is expressible as the conditional *LESS-SELECT*: *-less* → / N \_\_\_\_\_, requiring that, if *-less* is present, there must be a noun preceding it. If we simply take *LESS-SELECT*—which is effectively a subcategorization frame—to be a violable constraint in OT, it can then be assigned a high rank, given that *-less* is a level 2 affix (*crime-less*, etc.). In contrast, the similar conditional *AL-SELECT*: *-al* → / N \_\_\_\_\_ will be assigned relatively low rank, given that *-al* is level 1 (*crimin-al*, etc.). Since OO-FAITH equals *SELECT* + *gradient attraction*, this will result in OO-FAITH of two different ranks for level 1 and level 2 affixes (on *gradient attraction*, see below), consistent with Benua’s (1997) idea that affixes are subcategorized for OO-FAITH of different ranks. Similarly, the OO-FAITH imposed by the preposition *s* ‘with’ on its complement in Bulgarian can be directly related to the subcategorization frame for the preposition, *s-SELECT*: *s* → / \_\_\_\_\_ NP, evidently high-ranked, making the resulting OO-FAITH dominant over \*[?]. In the case of Russian [?]Amerikals Amerikoj ‘America/with America’, which, as Rubach (p. 289) notes, contrasts minimally with its Bulgarian counterpart in lacking the overapplication of ?-epenthesis, *s-SELECT* can be taken to be lower-ranked, resulting in OO-FAITH now being dominated by \*[?] (hence reversing the rank in (1b), while \*[?] remains dominated by *ONSET*, just as in (1a)).

As I argue in Burzio 1997, to appear a,b, larger units correspond to larger class sizes. For example, a preposition like *with*, which combines with NPs, produces a larger class of structures (PPs) than does an affix like *-al*, which combines with Ns (producing adjectives). The reason is

simply that there are more NPs than there are Ns, each N yielding multiple NPs when it itself combines with other constituents. Hence, the same combinatorial properties that yield larger units from smaller ones, also yield larger classes. Now the fact that there are more instances of *with*-NP than there are of *N-al* will—as desired—result in a higher-ranked conditional, *with* → / \_\_\_\_ NP, than the corresponding conditional *-al* → / N \_\_\_\_ on the assumption that the strength of such conditionals simply accrues over experience. Since there are more instances of *with*-NP than of *N-al* in the experienced data, the former conditional will acquire greater strength/higher rank than the latter. Since such SELECT conditionals are one of the bases of OO-FAITH, the latter constraints will then be correctly expected to be higher-ranked at phrasal levels than at the word level.<sup>3</sup> In addition to their rank, the conditionals themselves can be directly linked with experience on the hypothesis I advance in Burzio, to appear a,b, that mental representations of linguistic forms are sets of entailments, each entailment expressing dependency of some aspect of the representation on some other, as if any cooccurrence were mentally registered as necessary. On this view a conditional such as *-al* → / N \_\_\_\_ is an “emergent” property of the lexicon, due to the collective effect of all representations of the type *N-al* (see Burzio, to appear a,b, for more exact motivation). In sum, on the proposed view output-to-output faithfulness is motivated independently (aside from gradient attraction, discussed below) as just a specific interpretation of the type of selectional dependencies that are generally referred to as subcategorization specifications, and which any theory needs to express in some form. The present innovation is in taking them to be violable constraints, and in taking their rank to be driven by the frequency with which they are instantiated.

Reliance on derivations in characterizing word-to-word relations is undercut by two further considerations, one of which will lead us to the other basis of OO-FAITH: gradient attraction. This consideration is the simultaneous existence of multiple word-to-word relations. This claim is advanced independently in Burzio 1998 and Steriade 1999, on the basis of examples like (4a–b), respectively.

(4) a. *Italian*

vínc-ere	vín-t-o	vínc-it-óre
‘win (INF)’	‘won (PART)’	‘winner’

b. *French (dialect)*

[grɔs]	[gro]	[groz] arbre
‘large (FEM)’	‘large (MASC)’	‘large (MASC) tree’

In (4a) the agentive noun in *-óre* reveals a relationship with the participle *vín-to* by way of its participial affix *-it-*, related to the syncopated allomorph *-t-* of *vín-to*. This relationship with the

<sup>3</sup> The notion that OO-FAITH more strongly inhibits phonological alternations at phrasal levels than at word levels may seem challenged by languages that have extensive sandhi phenomena, such as Sanskrit (as one reviewer points out). I am not in a position to address such cases here. Note, however, that the existence of a generalization along the lines suggested in the text is confirmed by the independent existence of the Strong Domain Hypothesis once proposed in the context of Lexical Phonology (Kiparsky 1985). According to that hypothesis, no new phonological rules are permitted to be introduced at higher levels of compositionality. Rather, rules can only be discontinued at those levels, attesting to the increasing paucity of phonological alternations at higher levels.

participle is the normal one for *-óre* nouns in Italian (cf. *adatt-at-ol/adatt-at-ore* ‘adapted/adapter’, etc.). However, in (4a) the noun also reveals a relationship with the infinitive *vincere*, by sharing its consonant *c* [č], which the participle lacks. The case in (4b), from one dialect of French, is similar in that the liaison form [groz] shares the vowel of the masculine citation form, but the final consonant of the feminine form, aside from voicing (see Burzio 1999, Steriade 1999 for exact analyses). The existence of such multiple dependencies is expressible—in the form of multiple application of OO-FAITH—in the parallel system advocated here, but not by serial derivations. For instance, in (4a) the *-ore* noun could not be derived by calculating the structure of the inner participle first in cyclic fashion, since then the resurfacing of the underlying sequence /ci/ would be inexpressible ([vin-t]-o → [vin??-t]-ore). On the other hand, deriving the *-ore* noun from underlying material would miss important similarities between the participle and the noun and is therefore also excluded. As noted in Burzio 1998, syncope is variable in both the participle and the nouns, whence, alongside (4a), *tin-t-oltin-t-ore* ‘dyed/dyer’ (syncope in both), *aggred-it-ol/aggres-s-ore* ‘attacked/attacker’ (syncope in the noun only). In addition, syncope results in either an *-s-* or a *-t-* allomorph (from nonsyncopeated *-it-*), variably and unpredictably. Nonetheless, whenever both participle and noun syncope, they never fail to exhibit the same allomorph, as in *opprim-ere/oppres-s-oloppres-s-ore* ‘oppress/oppressed/oppressor’ (*-s-* allomorph) versus *redim-ere/reden-t-ol/reden-t-ore* ‘redeem/redeemed/redeemer’. Hence, at least for syncopeated nouns, the derivation would have to be cyclic. This is falsified, however, by the noted *aggred-it-ol/aggres-s-ore* case, with syncope in the noun only. Hence, there are no viable derivational routes for capturing the observed multiple dependencies (for more specific discussion, see Burzio 1998).

The surface-to-surface approach to (4) will of course raise the question of how an output-to-output faithfulness relation could exist between the *-ore* noun and the infinitive in (4a) if in fact *-ore* attaches to participles, and more generally how multiple correspondences arise on that approach. I argue in Burzio, to appear a,b, that such relations result from a general effect, *gradient attraction* (GA), by which similarity between representations generates pressure for further similarity. GA is the second component of OO-FAITH anticipated above, and, like SELECT, it results from the hypothesis that representations are sets of entailments. Hence, SELECT and GA are not independent provisions of the OO-FAITH machinery, but ultimately result from the same set of primitives (see Burzio, to appear a,b). In addition to these conceptual underpinnings, GA is independently motivated by the fact that, even for the same affixational class, faithfulness effects defy fixed ranking, instead being either stronger or weaker depending on overall similarity between trigger and target. For instance, in *remé:di-able* the constraint responsible for *CiV* lengthening (as in *CanadalCana:dian*, etc.) must outrank OO-FAITH relative to the base *remedy*, but the opposite must be true in *lévi-able* (short V) from *lévy*. The paradox is resolved by noting that *remedi-able* differs from its base verb in stress, a difference that weakens the ‘attraction’/OO-FAITH, while *leviable* has the same stress as *levy* and is thus subject to greater attraction/OO-FAITH (for further examples, and consequences, see Burzio 2000b, to appear a,b). Now in *vinc-it-óre* (4a), the stem *vinc-* will be subject to GA from both participle and infinitive, given its similarity to both in sound and meaning, but the participle will exert the greater attraction/higher-

ranked OO-FAITH because it is also ‘‘subcategorized for’’ by the affix *-óre* and hence also benefits from SELECT, the other component of OO-FAITH. As argued in Burzio 1998, one of the ways in which the participle exerts its influence/attraction is accentual: OO-FAITH<sub>stress</sub>, demanding that the stem in *vinc-it-óre* carry the same stress as *vín-t-o*. Given the stress on *-óre*, this requires the presence of an intervening syllable (to avoid a stress clash), and that is where the infinitive—the weaker attractor—comes in, lending needed segmental material (again, see Burzio 1998 for full discussion). Somewhat similarly, in (4b) the liaison form is expected to utilize the masculine citation form as its primary base, though here this is not because of affixal effects, but simply because the liaison form is itself masculine (greater independent similarity entailing stronger attraction). At the same time high-ranked ONSET forces divergence from that base, giving the feminine form (a secondary base/attractor) the opportunity to exert its influence, again lending needed material (see Steriade 1999 for details).

The second argument against derivations, besides the one based on multiple correspondences, relates to the behavior of affixes, illustrated by examples such as these:

- |        |                     |  |                               |
|--------|---------------------|--|-------------------------------|
| (5) a. | propagánda/Américan | propa(gánd-is)t $\emptyset$                      | A(mérica)(n-íst $\emptyset$ ) |
| b.     | títan               | *(títan-i)c $\emptyset$ /ti(tán-ic $\emptyset$ ) | bar(bár-ic $\emptyset$ )      |

As argued in Burzio 1994a,b, the stress-neutrality of (level 1) affixes like *-ist* in (5a) is best accounted for by taking the ‘‘metrical consistency’’ of the stem (a specific form of OO-FAITH) to prevail over that of the suffix, whose right edge can thus be parsed in one of two possible ways as indicated (similarly to what happens, for example, with the two independent items (*hónes*)t $\emptyset$ /ro(*búst*) $\emptyset$ ). In (5b), however, matters are reversed, as the metrical consistency (unique parse) of the suffix prevails over that of the stem, excluding stress-preserving \*(*títan-i*)c $\emptyset$ . In other words, the pressure for *títan-ic* to stress like *títan* is successfully countered here by the pressure for *-ic* to parse in the same way as it does elsewhere in the lexicon (e.g., as in *barbar-ic*). From the present perspective, this is just another GA/OO-FAITH effect: same segments, same meaning entail the same metrical parse. In a derivational system relying solely on IO-FAITH, the metrical invariance of *-ic* or any other stem-restressing suffix must be attributed to properties of the input. This solution is incorrect, however. As argued in Burzio 1994b:sec. 8.4, the divide between restressing and stress-preserving Latinate suffixes is predictable from output properties. The restressing suffixes like *-ic* and *-al* are those that do not provide enough material for a full metrical foot. Thus, unlike in *A(mérica)(n-íst $\emptyset$ )*, where both feet are well formed, thus allowing preservation of stem main stress, in (e.g.) *\*Na(póleo)(n-ic $\emptyset$ )* the final foot would be ill formed. Rather, well-formedness requires one more syllable, as in *Na(pòle)(ón-ic $\emptyset$ )*; but the resulting foot calls for main stress, with consequent stress shift from *Napóleon*. The generalization is that suffixes that are necessarily stem-restressing under some circumstances choose to avoid *their own* allomorphy by parsing in a fixed manner as in (5b). Hence, in the competition between stem and suffix OO-FAITH, the winner is whichever form of OO-FAITH is the more productive. While the problem of tying rank of OO-FAITH to productivity may be complex, requiring calculations over lexicons, the input approach is incorrect, because it predicts no generalization. Note also in this regard that the distinction between *-ic $\emptyset$*  with fixed presuffixal stress and other suffixes like *-al* or *-ous* with penultimate/antepenultimate stress (e.g., *bar(bár-ic $\emptyset$ )* versus (*bárbar-ou*)s $\emptyset$ ) is also not a matter

of input specifications. Rather, as argued in Chomsky and Halle 1968, the special metrical properties of *-ic* are best attributed to its massive alternation with *-ical*, as in *aca(dém-ic∅)/aca(dém-ica)l*. As argued in Burzio 1994b, the special parse of *-ic∅* is the only option available to satisfy the expected stem OO-FAITH over such pairs. The fact that the same parse obtains in the absence of an *-ical* variant (*Napole(ón-ic∅)??Napole(ón-ica)l*) then requires OO-FAITH over different instances of *-ic*. In sum, as a matter of factual description, affixes tend to metrical invariance just as much as stems. Derivational approaches introduce an unwarranted asymmetry: the invariance of stems is the result of cyclic derivations, while the invariance of affixes is the result of input prespecification. This conclusion is incorrect, because the exact modalities by which affixal invariance obtains are predictable from comparing outputs.

Recapping this section: Rubach admits that the facts in (1) provide no argument for derivations over output-to-output faithfulness. Other facts, however, support output-to-output faithfulness over derivations.<sup>4</sup> In particular:

- There is a morphophonological generalization about English level 2 affixation that has a unitary output-to-output faithfulness account but results in an unexplained conspiracy under a derivational account.
- Any theory of phonology has to postulate some grammar-internal modulation across different morphological domains such as underived; level 1; level 2; phrases. However, the failure to distinguish input-to-output faithfulness from output-to-output faithfulness inherent in Rubach's system yields a nonmonotonic fall-rise distribution in the ranking of IO-FAITH constraints distributed over what would otherwise seem a natural progression.
- Output-to-output faithfulness can be enforced simultaneously from multiple bases. It also applies across different instantiations of the same affixal material. In Burzio, to appear a,b, I lay out a framework in which the scope of output-to-output faithfulness can be defined in terms of GA and the selectional properties of affixes.<sup>5</sup> In contrast, neither effect is expressible by derivations.

<sup>4</sup> Rubach (p. 303) also advances the following argument against output-to-output faithfulness: Given an alternation like Bulgarian *obet/obed-i* 'dinner-SG/PL' due to final devoicing, output-to-output faithfulness predicts that a language may come to level all such alternations as in *obet\*obet-i*, which 'is unheard of.' Such leveling would result from the ranking IDENT<sub>OO</sub>([voice]), FINAL-DEVOICING >> IDENT<sub>IO</sub>([voice]).

This argument is successfully challenged by the following considerations: (a) It is not clear why such a situation in which final devoicing overapplies would be 'unheard of' given its parallelism to the *meteR/meteR-ing* case reviewed in the text, which requires a ranking just like the one Rubach argues against. A similar case is that of younger-generation Korean discussed in Kenstowicz 1996, in which the alternation *kap/kaps-i* 'price (CITATION FORM/NOM)', due to underlying /kaps/ and final cluster simplification, is now being leveled to *kap/kap-i*. (b) More importantly, Rubach's own system allows this situation to arise under the following grammar: final devoicing at level 1; plural affixation with undominated IDENT([voice]) at level 2.

There seems no way to exclude *obet\*obet-i* in the 'language change' scenario that Rubach considers, either. Whichever generation first drops final devoicing will have a choice of basing its underlying representation either on the old singular *obet*, producing exactly the unwanted *obet\*obet-i*, or on the old plural *obed-i*, producing *\*obed/obed-i*. There is no reason to believe that learners privilege plurals in setting up underlying representations.

<sup>5</sup> That framework also captures the noted 'base priority' effect; accounting for this effect is therefore not an advantage for derivations. More specifically, it captures the stronger tendency for morphologically simpler forms to influence their derivatives, without however categorically excluding 'back-copying' as in the previously noted *académicallacadémic* case (see Burzio, to appear a,b).

### 3 Opacity Cases

If conspiracies are a textbook argument against derivations, counterfeeding/counterbleeding effects have been an argument *for* them (Kenstowicz 1994:chaps. 2, 3). Rubach (pp. 280–283) uses derivations in the following two cases of phonological opacity in Slovak:

(6) <i>Opaque palatalization</i>		<i>Critical ranking</i>
a. Level 1	Palatalization: /pan-æ/ → pañ-æ	IDENT([-back]) >>> æ-BACKING
b. Level 3	æ-backing: pañ-æ → pañ-a (*pan-a)	æ-BACKING >>> IDENT([-back])
(7) <i>Opaque diphthongization</i>		<i>Critical ranking</i>
a. Level 2	Diphthongization: /ræ:sa/ → ryæsa	IDENT([-back]) >>> æ-BACKING
b. Level 3	æ-backing: ryæsa → ryasa (*rwasa)	æ-BACKING >>> IDENT([-back])

Rubach's use of level 1 for the case in (7) will be discussed below.<sup>6</sup> In (6) æ-backing counterbleeds palatalization, which otherwise occurs only before front vowels. In (7) æ-backing counterfeeds diphthongization in the sense that the back vowel *a* would be expected to select a back rather than a front glide. It also counterbleeds diphthongization altogether in that long [a:] does not diphthongize in Slovak. In Rubach's analysis it is therefore necessary in both cases that an underlying /æ/ be changed to [a] only late in the derivation. This is done by the "last-minute" change in the critical ranking indicated in (6) and (7). The question is, how can a last-minute effect be captured nonderivationally? Here again Rubach acknowledges the existence of a nonderivational alternative, in the form of McCarthy's (1999) sympathy theory. As Rubach notes, this theory permits the nonderivational identification, among all the candidates, of a "sympathetic" candidate that effectively has next-to-the-last-minute properties. For the case in (6) this candidate is *pañ-æ* and for the case in (7) it is *ryæsa*, identical to the input of Rubach's level 3. In McCarthy's system the correct output forms *pañ-a* and *ryasa* then result by assuming that faithfulness to this candidate appropriately dominates faithfulness to the input, thus ensuring a palatalized *ñ* in (6) and a diphthong in (7). Assuming further that æ-backing is undominated will ensure the correct output [a] rather than [æ] in both cases.<sup>7</sup>

<sup>6</sup> In (6) I have deliberately changed level 2 of Rubach's (20) to level 3 to amend what seems to be an inconsistency. If (6b) were level 2, its associated ranking would contradict that of (7a).

<sup>7</sup> Rubach (sec. 9.2) also considers and rejects a MAX(FEATURE) analysis of (6) and (7). Such an analysis would view the palatalization in (6) and presumably the frontness of the glide in (7) as preservation of the input feature [-back] present in underlying [æ], mirroring a proposal in Lombardi 1998. While the value of such an analysis of (6)–(7) may be limited since it would not generalize to other cases of opacity, Rubach's argument against it is dubious. He suggests that, on this analysis, /tel-æ/ should give \*[t'el-a] with nonpalatal / rather than actual [t'el'-a] 'calf' because the independent presence of [-back] in both [t'] and [e] should suffice to satisfy MAX([-back]). Surely this conclusion must be rejected, since features are not free to metathesize this way. Consider diphthong neutralization, as for instance in French diachrony: *fl[ew]r* → *fl[ø]r* 'flower'. Here the [+round] of the former glide is picked up by the vowel, thus motivating MAX([+round]). But no one doubts that this will happen regardless of the presence of some other [+round] in the string (which would already satisfy MAX([+round]) on Rubach's assumption): for example, *doul[ew]r* → *doul[ø]r* 'pain'. Hence, features, like segments, must be constrained by some form of contiguity. Rubach himself (fn. 39) notes the mirror image phenomenon in the Polish pronunciation of French [ü] as [yu], where the glide must be the effect of MAX([-back]) pertaining to the input vowel. Here again one can safely assume that the process will not fail in the presence of some other [-back] in the string.

Given Rubach's admission that there are nonderivational alternatives to both sets of cases discussed so far, at this point his argument for derivations can only rest on the disparate character of the output-to-output faithfulness and sympathy theories. However, such an argument is difficult to uphold, because Rubach's own analyses of the two sets of cases are rather unrelated to one another. In the case in (1) the two levels correspond to words and phrases as noted. But this is not true of the cases in (6) and (7). In particular, although applying "late," æ-backing appears to apply not at the phrase level, but "earlier." Since æ-backing in Slovak applies only after nonlabials, its application at the phrase level would predict word-initial [a]/[æ] alternations controlled by a preceding word-final segment. To the best of my knowledge, such alternations do not exist. Hence, the levels needed for (1) are unrelated to those needed for (6) and (7), leaving only the serial character of Rubach's solution to unify them. One could as easily claim that output-to-output faithfulness and sympathy are similarly unified by their common parallel character, leaving no argument in favor of Rubach's solutions.

More importantly, however, there exists a parallel account of opacity, which is well motivated, is superior to derivations, and can be naturally related to output-to-output faithfulness. This is the theory of targeted constraints developed by Wilson (2000, to appear). Wilson addresses the problem of opacity in OT by way of another problem, which he finds to be related. He notes that certain types of contextual markedness constraints of the form "Avoid X in context Y" are satisfied only by manipulating X, and never Y. Specifically, while one could say that it is marked for one obstruent to directly precede another, the crosslinguistically attested cluster simplification consists only of deleting the first, not the second, as in the Diola-Fogny example *let-ku-jaw* → *lekujaw*/\**letujaw* 'they won't go'—this despite the fact that either outcome would equally satisfy the surface prohibition. Wilson accounts for this asymmetry by introducing a set of what he terms *targeted* constraints, which effectively condition the repair in the right way. Targeted constraints formalize the role of contrast in phonology, along the lines of the dispersion theory developed by Lindblom (1986) and Flemming (1995), and they build on Steriade's (1994, 1997) claim that the positions subject to neutralization of segmental contrasts are those that limit the expression of perceptual cues relevant to the contrast. Targeted constraints penalize weak contrasts, while expressing the fact that weakness of contrast favors neutralization, as exemplified in (8), which yields the correct cluster simplification in the Diola-Fogny case.<sup>8</sup>

(8) →\**WEAK-C*

- a. Avoid an obstruent in preobstruent position.
- b. Candidate set: candidates that differ exactly by the weak element

This constraint expresses the fact that, owing to weakness of the relevant perceptual cues, an obstruent in preobstruent position effectively stands in a weak contrast with zero: (8a). By restricting the candidate set in the manner of (8b), it also expresses the fact that zero is a better alternative to the latter obstruent, and that nothing else is a better alternative for this particular constraint (other candidates, with more varied structural properties, are irrelevant). Diola-Fogny cluster

<sup>8</sup> The specific notation and some of the examples used in this discussion differ slightly (and inconsequentially) from Wilson's (2000).

simplification can now proceed as in (9), where I ignore epenthetic candidates (violating DEP) to simplify discussion.

(9)	let-ku-jaw	→*WEAK-C	DEP	MAX	*VELAR	*CORONAL
a.	letkujaw	*		a>b, c	*	*
b. <sup>ERASE</sup>	lekujaw	b>a		*	*	b>a, c
c.	letujaw	—		*	c>b, a	*
	Harmonic ordering	b>a	same	b>a>c	same	same

In (9) the asterisks indicate constraint violation as in standard OT notation. The cells unoccupied by asterisks are used for the harmonic ordering of candidates given by each constraint. For constraints other than the targeted one, this information is redundant with the asterisks. For example, the notation ‘‘a > b, c’’ under MAX indicates that candidate (9a) is better than candidates (9b) and (9c) according to MAX. The reason for this is that (9a) does not violate MAX, unlike (9b) and (9c) (as shown by the asterisks). The targeted constraint →\*WEAK-C, however, only compares candidate (9a), which violates it, with candidate (9b), in which the weak C has been removed, as required by (8b). Candidate (9c), in which something other than the weak C has been removed, is irrelevant on this approach. The bottom row in (9) then tallies the overall harmonic ordering of candidates, from left to right, prioritizing the higher-ranked constraints (i.e., conflicts are resolved by rank, as usual). On this analysis, candidate (9b) is correctly selected as the most harmonic and hence the winner (final harmonic ordering b > a > c). In contrast, as the distribution of asterisks clearly shows, standard OT would incorrectly pick (9c), so long as [coronal] is less marked than other places of articulation as expressed in the tableau and as generally assumed. The crucial factor in the correct selection is thus the exclusion of candidate (9c) by the targeted constraint.

Note that the cluster simplification problem, while unsolvable within standard OT, would appear to have a solution within a derivational theory of syllabification such as that of Itô (1986). In that framework one can postulate a syllabification stage banning medial codas, yielding *le<t>.ku.jaw*, followed by ‘‘stray erasure,’’ yielding the correct output. This analysis can presumably be reproduced in any framework that can duplicate Itô’s two stages, including DOT.<sup>9</sup> The derivational route seems precluded, however, by Steriade’s (1994, 1997) arguments that the contexts of neutralization are not always correctly identified by notions of traditional syllable theory such as ‘‘coda.’’ Rather, she argues that proper perceptual cuing is the relevant factor, such that a coda would be properly cued—on a par with an onset—if followed by a sonorant. If such arguments are accepted, as they are here or in Wilson 2000, to appear, a derivational account is excluded by the fact that perceptual cues are plainly properties of outputs, not of intermediate

<sup>9</sup> It can in fact be reproduced even in the original formulation of OT (Prince and Smolensky 1993), where faithfulness was defined in terms of PARSE and FILL, as shown in Zec 1995.

representations. Indeed, in Wilson's analysis the bound-for-neutralization candidate (9a) is a hypothetical output, not an intermediate representation in a derivation.<sup>10</sup>

Returning to opacity: Wilson shows further that targeted constraints can successfully deal with it as well. On that approach the [æ] → [a] neutralization of (6)–(7) would instantiate the work of a targeted constraint like (10), parallel to (8).

(10) →\*WEAK-æ

- a. Avoid [–back] with [+syllabic, +low], except after labials.
- b. Candidate set: candidates that differ exactly by the weak element

This constraint is meant to express the fact that, except after labials, the [æ]/[a] contrast is a weak one in Slovak—a situation that is less harmonic than neutralization of the contrast. On this view—just as in Rubach's analysis—the reason why [æ] neutralizes to [a] rather than vice versa must be that [æ] is more marked than [a] in the traditional OT sense of markedness, which is still required in the present system along with the “weak contrast” notion of markedness. Similarly, the deleted C in (9) must be more marked than its zero alternative. One might ask for evidence that [æ]/[a] is indeed a weak contrast, but that evidence is less than necessary in the present context. The present approach is based on the proposition, advanced by Steriade and Wilson and motivated independently, that neutralization occurs under weakness of contrast. Invoking it in the present case is a legitimate use of independent assumptions. It is parallel to Rubach's use of derivations, which are a general resource of his framework. Evidence may be discovered that [æ]/[a] is in fact *not* a weak contrast, falsifying the present analysis. Similarly, evidence may be discovered that Rubach's specific derivation is untenable. Until that point the two approaches seem comparably cogent.

The constraint in (10) will now give rise to the opaque interaction of (6), as in (11).

(11)

	pan-æ	→*WEAK-æ	IDENT([–back])	PALATN.	IDENT([–anterior])
a.	pan-æ	*	a>c, d	*	a>d
b.	pañ-æ	*	b>c, d	b>a	b>d
c.	pan-a	c>a	*	c>a	c>d
d. $\text{[ɛ̃]}$	pañ-a	d>b	*	d>a	*
Harmonic ordering		c>a, d>b	d>b>c>a <sup>11</sup>		same

<sup>10</sup> It is sometimes suggested that the type of phonetic grounding expressed in Wilson's analysis obtains only diachronically. The loss of the *t* in Diola-Fogny's *le(t)kujaw*, while indeed due to perceptual factors, would be the result of diachronic reanalysis. This view constitutes a weaker theoretical position, since it treats the relevant allomorphic variation as synchronically arbitrary. I see no need to abandon the stronger position, at the moment. One could equally well suggest that the stages of derivations like Itô's are to be interpreted only diachronically. The same objection would apply.

<sup>11</sup> IDENT([–back]) gives ambiguity between c>a>d>b and d>b>c>a, but PALATALIZATION selects the latter ordering.

Given (10b), the only harmonic orderings established by the targeted constraint in (11) will be  $c > a$  and  $d > b$  (i.e., here there are two candidate sets:  $\{c, a\}$ ,  $\{d, b\}$ ). While the overall harmonic ordering of candidates can be tallied in the bottom row step-by-step as before, we can more simply note that the last three constraints together favor the transparent, palatalized candidate (11b). The undominated targeted constraint can then be viewed as simply reasserting (11d) over (11b), causing it to win instead. As before, the success of the analysis hinges on the targeted constraint failing to make certain comparisons, specifically here between candidates (11b) and (11c). In standard OT, which makes all the comparisons, the transparent, nonpalatalized candidate (11c) would win, incorrectly, as is easy to see.

Calculation of the other opaque case,  $/r\text{æ}:sa/ \rightarrow [ryasa]$  in (7), would be similar and is left as an exercise for the reader.<sup>12</sup>

While Wilson's theory shares formal similarities with McCarthy's sympathy theory, it has certain empirical advantages over it. One is that it is independently motivated by the cluster simplification problem discussed above and by the need to formalize the critical notion of weak contrast. Another is that it is relatable to the output-to-output faithfulness theory, which is also independently motivated. I argue in Burzio 2000b (though in slightly different terms) that the critical characteristic of targeted constraints—the restriction on the candidate set, as in (8b) and (10b)—is essentially a restatement of the gradient attraction (GA) effect discussed above. The reason is that saying that a representation A attracts a similar representation A' (GA) is indeed equivalent to saying that harmony would improve if A' were to be replaced by A as in the pairwise comparisons made by targeted constraints. Unlike candidates (11b,d),  $[pañ\text{-}\text{æ}, pañ\text{-}a]$ , candidates (11b,c),  $[pañ\text{-}\text{æ}, pañ\text{-}a]$ , are not relatable by GA, since attraction of  $[\text{æ}]$  by  $[a]$  should have no effect on the rest of the string.

Attraction under a high degree of similarity, which thus seems to capture the nature of segmental neutralizations, is also detectable in patterns of allomorphy, as shown by the pair in (12) mentioned earlier.

- (12) a.  $lévy \rightarrow lévi\text{-}able$  (OO-FAITH  $\gg$  *CiV*-LENGTHENING)  
 b.  $rémedy \rightarrow remé:di\text{-}able$  (*CiV*-LENGTHENING  $\gg$  OO-FAITH)

These cases give rise to an apparent ranking paradox, with OO-FAITH taking on two contradictory ranks. The generalization underlying this and several other similar paradoxes examined in Burzio, to appear b, is that the dominated version of OO-FAITH is always found when base and derivative already differ along some other independent dimension. In (12b) the relevant difference, not shared by (12a), is stress. It thus appears that, just as with segmental contrast, with allomorphy proximity promotes neutralization. The apparent reranking of OO-FAITH in (12) follows from the fact that one of its main components, GA, varies in strength, while the other component, SELECT, enforced by the suffix, remains constant. One could see the high-ranked OO-FAITH/GA of (12a)

<sup>12</sup> Note, however, that \*WEAK- $\text{æ}$  will have to refer to short  $[\text{æ}]$  only, as reference to  $[\text{æ}:]$  would allow  $*ra:sa$  to win. This is consistent with the reasoning in the text: vowel length enhances contrast; hence, only short  $[\text{æ}]$  and  $[a]$  contrast weakly.

as in fact equivalent to a targeted constraint \*WEAK-LENGTH that bans the weak contrast *lévy/lé:vi-*, two structures differing only by the length of one vowel, while the corresponding contrast *remedy/remé:di-* is tolerated, because the two structures differ in stress as well and thus fail to occur in the same candidate set: (8b), (10b). Conversely, one could as well interpret the pressure to neutralize weak segmental contrasts as a form of OO-FAITH holding between close—mentally instantiated—segmental alternatives: [t] and zero in (9), and [æ] and [a] in (11). As noted above, the GA effect (which describes pressure for identity as being inversely related to the multidimensional distance between two representations) is derivable from a specific assumption on the nature of mental representations: that they constitute sets of entailments, a point discussed in Burzio, to appear a,b.

Once this relationship is established between output-to-output faithfulness and the targeted constraints, it is no longer surprising that both can account for cases of opacity, like (1) and (11), respectively. That account is essentially unitary from the present point of view.

In sum, the facts discussed so far would provide no real argument for derivations even if McCarthy's (1999) sympathy theory was adopted alongside output-to-output faithfulness. These two extensions of OT may seem unrelated, but so do the two derivational extensions proposed by Rubach. On the other hand, as also argued in Burzio 2000b, Wilson's (2000, to appear) approach to opacity finds a direct conceptual link with output-to-output faithfulness within the perspective developed in Burzio, to appear a,b. It also benefits from the independent motivation provided by the cluster simplification asymmetry of (9), which resists derivational treatments.

#### 4 Other Cases

The above discussion has defused all of the potential arguments for derivations in Rubach's analyses—essentially the ones that rely on opacity effects. In this section I review two further cases that Rubach invokes as arguments for derivations, but where the use of derivations seems spurious.

One such case concerns the contrast in standard Slovak between *ryasa* (← /ræ:sa/) 'cassock', discussed in the previous section, and *dialekt* 'dialect'—specifically, the contrast between the diphthong [ya] and the bisyllabic sequence [ia]. Rubach (pp. 280–282) proposes the derivational account in (13).

(13)				<i>Critical ranking</i>	
	UR	/ræ:sa/	/dialekt/		
	a. Level 1	Syllabification:	ræ:.sa	di.a.lekt	NO-DIPH >> ONSET
	b. Level 2	Diphthongization:	ryæ.sa	di.a.lekt	IDENT(Nuc) >> ONSET >> NO-DIPH
					*[æ:] >> NO-DIPH
	c. Level 3	æ-backing:	rya.sa	di.a.lekt	See (7b)

It is clear that no derivation would be needed to account just for *di.a.lekt*. So long as the ranking NO-DIPH >> ONSET holds just as in Rubach's level 1 grammar, the input /dialekt/ will give the correct output. The derivation is only called for by the following incorrect assumption: If the

ranking NO-DIPH  $\gg$  ONSET held throughout, then level 2 *ryæ.sa* (with a diphthong, satisfying ONSET) would become *ri.æ.sa* (with a hiatus, violating ONSET), yielding surface *\*ri.a.sa*. On this assumption, that ranking has to be reversed at level 2 to maintain *ryæ.sa*, while *di.a.lekt* is also kept syllabically frozen by IDENT(Nuc). The assumption is incorrect because long vowels do not choose to turn into diphthongs rather than bivocalic sequences, because compelled to do so by ONSET. If they did, then, for example, Italian [fwɔ.ko] ‘fire’, diphthongized from Latin [fɔ.ku], might just as well have become \*[fu.tɔ.ko] instead—an alternative satisfaction of ONSET, but an unimaginable outcome. Rather, long vowels diphthongize because of incompatibility between their prosodic length and some specific feature (often [–ATR]). Hence, the vowel with that specific feature stays short, while prosodic length is achieved by adding the glide, as shown in (14).

(14)

	fɔko	STRESS-TO-WEIGHT	*[ɔ:]	DEP	NO-DIPH
a.	fɔ.ko	*			
b.	fu.(t)ɔ.ko	*		*	
c.	fɔ:.ko		*		
d.	fwɔ.ko			*	*

In (14) prosodic length is imposed by a STRESS-TO-WEIGHT constraint (Burzio 1994b:sec. 2.4). Diphthongization occurs under prosodic equivalence, both [ɔ:] and [wɔ] satisfying STRESS-TO-WEIGHT. The breakup of a long vowel into a bivocalic sequence would violate that equivalence—a crucial point that Rubach himself notes in his footnote 15, yet does not take into account. Thus, while Proto-Romance lengthening without diphthongization would give \*[fɔ:.ko] (14c), there would be no reason to choose either variant of (14b) \*[fu.ɔ.ko] or \*[fu.tɔ.ko] instead, since both violate the lengthening imposed by STRESS-TO-WEIGHT in addition to violating some form of DEP. Either structure would automatically lose to simpler [fɔ.ko] (14a), which only violates the lengthening. Hence, the short vowel option (14a) is guaranteed to ‘harmonically bound’ the bivocalic option (14b), ONSET playing no role whatever. This is true in the Slovak case as well, as shown in (15), where length is now imposed by a faithfulness constraint, since length is underlying in Slovak. (I ignore æ-backing here.)

(15)

	ræ:sa	IDENT(Length)	*[æ:]	DEP	NO-DIPH
a.	ræ.sa	*			
b.	ri.æ.sa	*		*	
c.	ræ:sa		*		
d.	ryæ.sa			*	*

As in (14), candidate (15b) *ri.æ.sa* from input *ræ:sa* directly loses to (15a) *ræ.sa* with a short vowel, since *ri.æ.sa* has a short vowel anyway (harmonic bounding). Hence, in addition to including a constraint banning length with certain features, the grammar of diphthongization must include a prosodic constraint requiring a long vocalic nucleus, dominating NO-DIPH. This is STRESS-TO-WEIGHT in (14) and IDENT(Length) in (15). The conclusion is that ONSET plays no role in diphthongization. Hence, there is no ranking paradox like the one in (13a–b). Hence, there is no need to rerank the constraints at different levels. Hence, there is no need for the levels.

The second case concerns the asymmetrical outcomes from the symmetrical inputs *Vu* and *uV* in standard Polish.

- (16) a.  $Vu \rightarrow Vw$  (p[aw]sa ‘pause’)  
 b.  $uV \rightarrow uwV$  (pap[uwa]s ‘Papuan’)

Rubach proposes the derivations in (17).

- (17) *Critical ranking*  
 UR /pau.sə/ /papu.as/  
 a. Level 1 Syllabification: paw.sa pa.pu.as \*ONSET([u]) >> ONSET  
 b. Level 2 Resyllabification: — pa.pu.was IDENT(Nuc), ONSET >> \*ONSET([u])

However, these derivations have no other function than stating an exception to the generalization that [w] in Polish cannot be an onset or part of an onset: namely, when [w] is adjacent to [u], as in [pa.pu.was]. Crucially, unlike in the opacity cases, this generalization is surface true. It can be expressed by means of the constraint in (18).<sup>13</sup>

- (18) \*INDEP([w])

The features of [w] must be dependent on an adjacent segment.

The critical cases are now simply calculated as in (19) and (20) under the given ranking, with no need for derivations.

(19)

	/papus/	ONSET	*ONSET([u])	*INDEP([w])
a.	pa.pu.as	*		
b.	pa.pwas		*	*
c.	pa.pwu.as	*	*	
d.	pa.pu.was		*	

<sup>13</sup> The constraint in (18) needs to be dominated by others, including the one(s) responsible for vocalizing “dark” /s, as in /skoła/ → [skowa] (Rubach 2000:292–295). No difficulty arises in this connection. Constraint (18) is also violated by the winner in (20).

(20)	/pausa/	ONSET	*ONSET([u])	*INDEP([w])
a.	pa.u.sa	*		
b.	pa.w.sa			*
c.	pa.wu.sa		*	
d.	pa.uw.sa	*		

In (19) violation of \*ONSET([u]) is forced by ONSET (epenthesis must be excluded by high-ranked ONSET, just as in Rubach's analysis). \*INDEP([w]) then chooses between the candidates satisfying ONSET, (19b) and (19d). In (20) realization of /u/ as [w] is optimal because it only violates bottom-ranked \*INDEP([w]), \*ONSET([u]) being satisfied by [w]'s parse as a coda—an option not available in (19). As Rubach notes, structures like (20c) (e.g., [xì.drà.wú.lik] 'plumber') are also allowed, however, in addition to (20b) (or [xì.dráw.lik]). Rubach's analysis of such cases in terms of input specification carries over to the present discussion straightforwardly: in [xì.drà.wú.lik] the [u] is syllabic (nuclear) in the input. The correct output is then guaranteed by the ranking IDENT(Nuc), ONSET >> \*ONSET([u]).

Note that the constraint in (18) also seems to have some independent justification, as it is reminiscent of the one in (21) needed for Italian.

(21) \*INDEP([−sonor])

The features of an obstruent *in coda position* must be dependent on an adjacent segment. (Latin *nocte* 'night', *optimu* 'optimal' → Italian *notte*, *ottimo*, etc.)

Within the licensing-by-cue approach proposed by Steriade (1997) and Wilson (2000), preobstruent codas are weak positions for other obstruents because they obscure perceptual cues to place and other features. This impasse is resolved by sharing features with a following onset, which is itself well cued, as reflected in (21). Similarly, one could take onsets to be weak positions for glides ([w] in particular) because of their weak closure, and it seems plausible that sharing features with an adjacent vowel would help perceptual identification. While (18) does not specifically refer to onset position, it could do so without affecting the discussion (the violation in (20b) would just be removed). Whatever the strength of this independent justification, it compares favorably with Rubach's derivational account, which will not carry over to (21).

## 5 Conclusion

Rubach (2000) proposes derivational analyses for a number of phonological alternations in Slavic. His main argument against nonderivational alternatives is that none of them can cover the full range of cases. That argument is undercut by the fact that his own derivational analyses do not constitute a unitary solution. It is further undercut by the fact that unity can be established within a nonderivational approach. I have divided the cases in point into three sets: those amenable to

an analysis in terms of output-to-output faithfulness; those analyzable in terms of Wilson's (2000, to appear) targeted constraints; and those that can be analyzed directly within simple OT. I have argued for a link between output-to-output faithfulness theory and targeted constraints. The link is provided by the fact that one basis for output-to-output faithfulness is similarity of representations—precisely the basis for targeted constraints (Burzio 2000b). This suggests that ‘attraction’ under proximity in multidimensional space is a general property of mental representations (Burzio, to appear a,b). In addition to their unitary basis, output-to-output faithfulness and targeted constraints each enjoy independent motivation that the derivational solutions lack. Unlike derivations, output-to-output faithfulness captures the generalization that those domains that exhibit limited phonological alternations are also morphologically more regular (e.g., English *crime-less*, lacking both vowel shortening and morphological idiosyncrasy, compared with *crimIN-al*). Unlike derivations, targeted constraints relate neutralization processes to the same considerations of perceptual distance that are needed to successfully characterize the structure of segmental inventories. Unlike reintroducing derivations, comparing outputs with ‘other’ representations, which both output-to-output faithfulness and targeted constraints perform, is a conceptually natural deployment of OT's parallel resources. The above discussion leads to the welcome conclusion that what seems conceptually natural is also empirically supported.

## References

- Benua, Laura. 1997. Transderivational identity: Phonological relations between words. Doctoral dissertation, University of Massachusetts, Amherst.
- Borowsky, Toni. 1993. On the word level. In *Phonetics and phonology 4: Studies in Lexical Phonology*, ed. Sharon Hargus and Ellen Kaisse, 199–234. San Diego, Calif.: Academic Press.
- Burzio, Luigi. 1994a. Metrical consistency. In *Language computations*, ed. Eric S. Ristad, 93–125. Providence, R.I.: American Mathematical Society.
- Burzio, Luigi. 1994b. *Principles of English stress*. Cambridge: Cambridge University Press.
- Burzio, Luigi. 1996. Surface constraints versus underlying representation. In *Current trends in phonology: Models and methods*, ed. Jacques Durand and Bernard Laks, 123–141. Salford, England: University of Salford Publications.
- Burzio, Luigi. 1997. Strength in numbers. In *University of Maryland working papers in linguistics 5: Selected phonology papers from H-OT 97*, 27–52. Department of Linguistics, University of Maryland, College Park.
- Burzio, Luigi. 1998. Multiple correspondence. *Lingua* 103:79–109.
- Burzio, Luigi. 2000a. Cycles, non-derived-environment blocking, and correspondence. In *Optimality Theory: Phonology, syntax, and acquisition*, ed. Joost Dekkers, Frank van der Leeuw, and Jeroen van de Weijer, 47–87. Oxford: Oxford University Press.
- Burzio, Luigi. 2000b. Segmental contrast meets output-to-output faithfulness. *The Linguistic Review* 17: 368–384.
- Burzio, Luigi. To appear a. Missing players: Phonology and the past-tense debate. *Lingua*. [Preliminary version in *University of Maryland working papers in linguistics 10: Proceedings of the Maryland Mayfest on Morphology 1999*, 73–112. Department of Linguistics, University of Maryland, College Park.]
- Burzio, Luigi. To appear b. Surface-to-surface morphology: When your representations turn into constraints.

- In *Many morphologies*, ed. Paul Boucher. Somerville, Mass.: Cascadilla Press. [Preliminary version: Rutgers Optimality Archive ROA-341-0999; <http://roa.rutgers.edu>.]
- Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.
- Flemming, Edward. 1995. Auditory representations in phonology. Doctoral dissertation, UCLA, Los Angeles, Calif.
- Itô, Junko. 1986. Syllable theory in prosodic phonology. Doctoral dissertation, University of Massachusetts, Amherst. [Published New York: Garland Press (1988).]
- Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.
- Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Cambridge, Mass.: Blackwell.
- Kenstowicz, Michael. 1996. Base identity and uniform exponence: Alternatives to cyclicity. In *Current trends in phonology: Models and methods*, ed. Jacques Durand and Bernard Laks, 363–393. Salford, England: University of Salford Publications.
- Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2:82–138.
- Kiparsky, Paul. 1998. Paradigm effects and opacity. Ms., Stanford University, Stanford, Calif.
- Lindblom, Björn. 1986. Phonetic universals in vowel systems. In *Experimental phonology*, ed. John J. Ohala and Jeri J. Jaeger, 13–44. New York: Academic Press.
- Lombardi, Linda. 1998. Evidence from Max feature constraints from Japanese. In *University of Maryland working papers in linguistics 7: Papers in phonology*, 41–62. Department of Linguistics, University of Maryland, College Park.
- McCarthy, John J. 1999. Sympathy and phonological opacity. *Phonology* 16:331–399.
- Prince, Alan, and Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Ms., Rutgers University, New Brunswick, N.J., and University of Colorado, Boulder.
- Rubach, Jerzy. 2000. Glide and glottal stop insertion in Slavic languages: A DOT analysis. *Linguistic Inquiry* 31:271–317.
- Steriade, Donca. 1994. Positional neutralization and the expression of contrast. Ms., UCLA, Los Angeles, Calif.
- Steriade, Donca. 1997. Phonetics in phonology: The case of laryngeal neutralization. Ms., UCLA, Los Angeles, Calif.
- Steriade, Donca. 1999. Lexical conservatism in French adjectival liaison. In *Formal perspectives on Romance linguistics*, ed. Barbara Bullock, Mark Authier, and Lisa Reed, 243–270. Amsterdam: John Benjamins.
- Wilson, Colin. 2000. Targeted constraints: An approach to contextual neutralization in Optimality Theory. Doctoral dissertation, Johns Hopkins University, Baltimore, Md.
- Wilson, Colin. To appear. Positional faithfulness, licenser avoidance, and harmonic ordering. *Phonology*.
- Zec, Draga. 1995. The role of moraic structure in the distribution of segments within syllables. In *Frontiers of phonology: Atoms, structures, derivations*, ed. Jacques Durand and Francis Katamba, 149–179. New York: Longman.

*Department of Cognitive Science*  
*Johns Hopkins University*  
*Baltimore, Maryland 21218*  
 burzio@jhu.edu