Elsewhere & Otherwise

Alan Prince
Rutgers University

In a contribution to *Glot International 1.9/10*, Morris Halle finds rather little to praise in recent work. But various misinterpretations and flaws of logic combine to nullify his qualms and, backandedly, suggest a rosier picture of current research directions.

1. Rule-Package Serialism

Halle's larger project is to defend the formal assumptions of Chomsky 1951/ Chomsky 1965/ Chomsky & Halle 1968, and in particular two basic ideas —

(1) Grammatical generalizations are packaged into re-write rules which fuse a Structural Description to a Structural Change;

(2) Such *rules* are extrinsically ordered with respect to each other.

Point (1) by itself excludes the many post-1968 linguistic theories which respond to the fundamental insight that the principles of form are scattered and lost in the Aspects/SPE rule-package device, and which emphasize the role of universal principles in grammar. (Major, early points of departure include Kisseberth 1970 and Chomsky 1977.)

A basic sense of the issues can be garnered from a look at Halle's one stated rule: "Shorten the head vowel of a branching foot." This brief imperative fairly bristles with unasked questions.

○ Why *shorten* rather than *lengthen*? indeed, why do anything at all?

○ Why *head* rather than *non*head ? indeed, why not affect both?

○ Why *branching* rather than *non*branching? indeed, why not operate on any foot?

Over the last decade or so, metrical theory has developed answers to exactly this kind of question, and recent formal work is much concerned with mechanisms for integrating such answers into grammars.

■Assume a strict bimoraic perspective[1] on trochaic rhythm, which disallows Heavy-Light feet. The cited mapping serves to eliminate an unparsed syllable, converting (H)L to (LL). Optimality Theoretically, this is an effect of PARSE-σ dominating the Faithfulness constraint that requires input-output identity of vowel-length.

■Assume a more syllabically-based perspective[2] on binarity, by which the foot (HL) is admitted. Then constraints on the rhythmic wellformedness of the various possible foot-shapes drive the breach of faithfulness.

Either way, the very same principles at work in the simple assignment of foot-structure — PARSE-σ, rhythmic harmony — are also effective in other prosodic processes. The questions raised by the rule-

---

[1] See e.g. Allen 1973; McCarthy & Prince 1986; Hayes 1986/1987,1995; Mester 1995.

[2] This is the view of early metrical theory, re-examined from a pre-OT optimization perspective in Prince 1990. See Churchyard 1991 for the first reformulation within OT.

package are answered in this way: the observed mapping, but none of the mentioned alternatives, brings the structure into greater accord with the principles of prosodic form. There is no separate formal machinery whirring away independent of the basic theory of prosodic structure.[3]

Such results are achieved in Optimality Theory, as in other approaches developed since the 1970's, by liberating the structural constraints from the parochial SC/SD rule-package. Halle is silent on the explanatory problems that beset the rule-package idea. This suggests that there may be an as-yet-unacknowledged watershed division in methodology underlying his position, of the kind that seems to occur in syntax every 5 to 10 years, having to do with the relative weight placed on explanation-from-principle vs. descriptive completeness. But unless the explanatory issues are explicitly addressed, it's not clear what there is to defend in the rule-package theory.

2. Elsewhere Logic

In aid of his project, Halle argues that an *Elsewhere Condition* on rule-interactions governs the relation between the vowel-shortening rule of English stated above and a vowel-lengthening rule of that language. (The same example is presented under the same rubric in Kenstowicz 1994:21.) Halle believes that Elsewhere-type interactions pose special problems for Optimality Theory, although grounds for this belief are not given, and are unlikely to be found. Observe that the Elsewhere Condition determines which one of two or more competing processes will apply in a given environment, blocking the expected derivational relationship between them. The aim, then, is to argue from the *failure* of serial derivation to the conclusion that serial derivation is necessary: the logic is not going to be straightforward. By contrast, Optimality Theory is thoroughly in the business of selecting which (of many) competing input-output maps will prevail and needs no Elsewhere Condition to guide it. As Prince & Smolensky observe, Elsewhere Condition disjunctivity is a sub-case of the general mode of constraint interaction in OT. And, of course, the OT analytical literature already contains numerous instances of special-case/general-case interactions that work as expected (e.g. Prince & Smolensky 1993:111; McCarthy & Prince 1995:31ff., and many others on complementary distribution, etc.).

Is it even clear that the logic of the relation between the Elsewhere Condition and serial rule ordering has been worked out sufficiently to support such an argument? The Kenstowicz/Halle example does not in fact provide much evidence for adjoining an Elsewhere Condition to extrinsic-ordering models of the SPE type: the relationship between the two rules is already obtained by the theory of serial ordering. Special lengthening is ordered *after* General shortening, and simply undoes the general shortening in its own narrow environment. This kind of effect, by which 'special' imposes itself on 'general' through ordinary rule-application, is predicted by ordering theories like SPE: there is no obvious need to re-predict it. In some versions of Lexical Phonology, even the derivational relation between the two English rules can be deduced from their other properties. On the basis of differential 'derived environment' behavior, Kiparsky (1982:44) places the lengthening rule in the postcyclic phonology, and the shortening rule ("Trisyllabic Shortening") in the lexical

---

[3] Observe that the Myers 1987 analysis, which Halle modifies, understands the shortening effect in terms of otherwise-motivated processes of the language, interacting with universal conditions; Myers seeks to eliminate the shortening rule entirely, not just to re-phrase it. For another perspective, see Burzio 1995. See Baković 1996 for analysis of the English lengthening-shortening system.

cycle, thereby determining the interaction by componential affiliation rather than extrinsic ordering. SPE, of course, uses straight ordering, cf. SPE:240-241, ex. (20) and SPE:242, ex. (23.IV).

Stepping back from the particularities, we note that ordering theories already possess two distinct mechanisms that achieve 'Elsewhere'-looking effects. One is overwriting, as just seen: a later rule can reverse the effects of an earlier one. The second is simple bleeding: an earlier rule can produce a structure that the later one does not apply to. For example, if metrical structure assignment operates under the Free Element Condition (Prince 1985; Halle & Kenstowicz 1991), then the application of a later rule of metrical structure assignment is blocked — bled — by the mere presence of the earlier-assigned structure. Along the same lines, feature-filling rules are blocked by the presence of specified features, which may be assigned by rule. If the structural descriptions of the relevant rules stand in the special-general relationship, then in both the overwriting and the bleeding regimes we will typically have a surface distribution of forms that can be described in 'elsewhere' language, although no Elsewhere Condition participates in their derivation. In view of this, it is surprising that no analyst has framed a *Panini's Theorem on Rule Ordering*, describing the conditions under which elsewhere effects follow from serial ordering; perhaps this omission is due to a general belief that rule ordering is a descriptive convenience, rather than a potential source of constraint. (It should be noted that Janda & Sandoval (1984) tabulate a number of cases where, they assert, rule-ordering solutions can handle Elsewhere Condition based analyses.)

A higher-level argument might still be attempted: rule-ordering theory plus the Elsewhere Condition, as accepted by Halle, demands that such rules can interact in only one way, whereas the basic unadorned rule-ordering theory predicts two possible modes of interaction, one for each of the two orders.[4] (For the sake of argument we gloss over the fact that, given known formulations of the EC, which are highly sensitive to details of rule-content, the EC can often be formally circumvented by careful rule-writing.[5]

---

[4] Ordering and disjunctivity are, of course, two quite different notions, and there is no a priori reason to assume within SPE serialism that the form of rules can have any effect on their ordering; rather the opposite. Thus it is notably odd in this context to find a formal relation between rule-statements fixing their ordering. But the version of the EC assumed by Halle, current since Kiparsky 1982:136-7, forces an ordering rather than presupposing one, and in this assimilates the *Proper Inclusion Precedence Principle* (Sanders 1970, 1974, Koutsoudas, Sanders, & Noll 1971/74), as Kiparsky notes. The original Elsewhere Condition formulation in Kiparsky 1973:94 begins "Two adjacent rules of the form..." retaining the SPE idea, obviously necessary for reduction via parentheses, that disjunctivity requires adjacency. Proper Inclusion Precedence was offered as one principle among several that determined the applicability of rules in a derivation, based on their form. Goldsmith (1984:36) observes the notional independence of the Elsewhere blocking relation and rule ordering, arguing that a *later* special rule will block an earlier-ordered general rule . Halle & Vergnaud (1982:81) dismiss Goldsmith's conception as unprecedented.

[5] In the case at hand, if the general shortening rule is written so as to apply to *long vowels*, then the SD of the lengthening rule, whether written to apply to short vowels or just vowels in general, no longer stands in the required substructure relationship, and the EC is not invoked. (Thanks to Eric Baković for bringing this point to my attention and discussing its significance.) Notice that the Evaluation Metric does not apply in this case to force elimination of the length specification in the SD of the shortening rule, because such specification is not redundant: it serves to disable the EC and thereby allows a different, nonequivalent grammar to be framed. Because of this unresolved formal quirk of definition, the EC add-on actually enriches descriptive capacity. We will argue, however, as if a more stable definition had been articulated.

The Elsewhere-advocate must then argue that, in cases like the one at hand, the unadorned serialist interaction 'Special precedes General without disjunction' is in fact impossible. More precisely, the argument must be that this interaction is impossible within a Lexical Phonological Level, the domain over which the Elsewhere Condition holds sway.

A single observation of type A, such as Halle aims to provide, cannot entirely persuade us that the universe lacks, or ought to lack, *not*-A. The force of the argument-from- one-data-point is further subverted by the character of the rule interaction. If lengthening (special) were to precede shortening (general) in a serial grammar, then in the simplest case the later rule would completely wipe out the effects of the earlier rule, leaving no long vowels produced by it. Absent other interactions to diagnose the hidden presence of the earlier rule, there would no reason at all for the rule-learner to posit such a rule in the first place. So the SPE-type theory predicts that getting two such rules in the same grammar, with the special-before-general order, will be possible only in very particular circumstances, sure to be rare. (There would have to be at least one derivational instant at which an underlying short vowel behaves as long in the peculiar environment of the lengthening rule, before winking back again to shortness by the general rule.[6]) The actual dispute over the treatment of such mutually-undoing pairs of rules, then, is between the straight rule-ordering theory, which predicts that one of the interactions is going to be *rare*, and the rule-ordering + EC qua Proper Inclusion Precedence theory, which predicts that one of the interactions is going to be *impossible* (within a given Lexical-Phonological component). On the face of it, then, the simple rule-ordering theory provides a complete account of the observed interactions of mutually-undoing rules: in the general-special order, the special rule undoes the general; but the opposite order will be rarely observed, since in a special-general sequencing of this type, 'general' will typically overwrite 'special' completely. The serialist can therefore take justifiable pride in predicting the interactions directly from the character of the rules, using only the most fundamental tools of the theory. There would not seem to be promising grounds here for a coarse-grained argument that the basic serial theory needs to be modified.

3. Elsewhere and Optimality Theory

Halle characterizes Prince & Smolensky 1993:108 as an attempt to reject the Elsewhere Condition on empirical grounds. He has it that those authors are trying to "set aside [the EC] on the basis of a few putative counterexamples" and answers that "one cannot conclude that the Elsewhere Condition is invalid, for there are many examples that support the Condition." The notion here is that Prince & Smolensky are thoughtlessly discarding "an empirical result of some importance."

The reality is distinctly otherwise. The cited passage argues for nothing more dramatic than the claim that the Elsewhere Condition holds only of cases where rules are incompatible in their effects, and not where the rules are identical in their effects. (These are the two classes of cases

---

[6] Kiparsky 1973:98-100 offers an interesting 'external evidence' argument from Rigvedic compositional practice in favor of the position that disjunctivity rather than seriality holds between special/general case rule-pairs: the extra representation provided by serial application ought to be, but is not, metrically detectable. Of course, various specialized assumptions about the relation between metrics and morphophonemics, and about the morphophenemics itself, are required for the argument to go through. See Howard 1975 for discussion. One could also imagine that empiricist squeamishness about long derivations would count as an argument against a strict serialist anti-Elsewhere position, to some minds (for discussion, see Pullum 1976).

identified in Kiparsky 1973:94; Kiparsky 1982:173fn. 2, already drops the identicality subcondition.) Optimality Theory, which adjudicates conflicts, gets the first class of cases as emergent from simple ranking and re-ranking of constraints, but can say nothing about the second; therefore, it is important to establish that the second class is correctly understood in different terms: i.e., that the 'identical effect' subclass really involves a form of incompatibility. It seems likely that the notion of 'identical effect' arose from construing 'effect' too narrowly, as nothing more than the Structural Change of a rule, rather than in terms of the broader structural effects to which constraints are now known to be sensitive. For example, a rule assigning final main stress and another rule assigning it penultimately have 'identical effect': they both assign main stress — but the results are clearly incompatible, when 'effect' is viewed from a product-oriented perspective. If the Prince & Smolensky argument holds overall, then Optimality Theory, which offers a general theory of prioritization, stands a good chance of accommodating the specific prioritizations of the Elsewhere type, without positing an Elsewhere Condition. In contrast, rule-ordering theory with an adjoined Elsewhere Condition forms a redundant and centaur-like composite: the supplementary theory of prioritization (Elsewhere) sits atop the theory of rule-ordering itself, which is already quite capable of modeling various prioritization effects, as has been noted above.

4. Faithfulness

Moving to the larger stage, Halle rejects the existence of faithfulness constraints, which are fundamental to Optimality Theory, in a single broad stroke: "the existence of phonology in every language shows that Faithfulness is at best an ineffective principle that might well be done without." This assertion is puzzling indeed. It is no argument against a theory of constraint violation, to observe that, in it, constraints are violated. More precisely, it is no argument against Optimality Theory to note that the constraints it predicts to be violable are in fact violated, and indeed, if one wishes to look a little further, violated in the way it predicts. The sense of Faithfulness in the context of Optimality Theory (and what other context do you find it in?) is that input-output disparity is minimized, not absent. This is a consequence of the 'minimal violation' principle that governs all constraints, not just those of the faithfulness families, and which is absolutely fundamental to the way the theory characterizes grammatical mappings.

5. Conclusion

Far from demonstrating the superiority of serial rule-package theories, Halle's discussion either fails to contact its putative subject matter (Prince & Smolensky on Elsewhere, Faithfulness) or fails to build an argument for his own positions (that Elsewhere effects blocking serial derivational relations show the need for rule-package serialism, that English shortening/ lengthening even displays an Elsewhere effect). Other wider claims — *e.g.* that rule-package serialism tells us "how speakers go from the neurological representation to the articulatory activity/acoustics" — seem so lacking in foundation (and so unconnected with anything that is known about the nervous system, articulation, or perception) as to resist detailed analysis. Ultimately, perhaps, it is the lack of any defense of the superficially appealing but deeply problematic rule-package idea that makes one wonder whether the place where SPE has pitched its mansion, Halle's slippery vantage, is a suitable ποῦ στῶ. The Hallean critique shows unambiguously, however, that there are still interesting things to be figured out about how conditions like Elsewhere fit onto complex re-writing systems, and in indirectly raising the difficult, classic problems, it commends to our attention recent approaches that grapple with them directly, and sometimes successfully.

Acknowledgment

References

Allen, W. S. 1973. *Accent and Rhythm; Prosodic Features of Latin and Greek : A Study in Theory and Reconstruction*. Cambridge University Press: Cambridge.

Baković, E. 1996. Elsewhere effects in Optimality Theory. Ms., Rutgers University: New Brunswick, NJ

Burzio, L. 1994. *Principles of English Stress*. Cambridge University Press: Cambridge.

Chomsky, N. 1951. The Morphophonemics of Modern Hebrew. MA Thesis, University of Pennsylvania: Philadelphia.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. MIT Press: Cambridge.

Chomsky, N. 1977. On Wh-Movement. In *Formal Syntax*, Peter Culicover, Thomas Wasow and Adrian Akmajian, eds.. Academic Press: New York, pp. 71-132.

Chomsky, N. and M. Halle.1968. *The Sound Pattern of English.* Harper & Row Publishers: New York.

Churchyard, H. 1991. Biblical Hebrew prosodic structure as a result of preference-ranked constraints. Ms., University of Texas: Austin.

Goldsmith, J. 1984. Tone and Accent in Tonga. In *Autosegmental Studies in Bantu Tone*, G.N. Clements and J. Goldsmith, eds. Foris Publications; Cinnaminson, pp. 19-51.

Halle, M. and M. Kenstowicz. 1991. The Free Element Condition and Cyclic vs. Noncyclic Stress. *Linguistic Inquiry* 22, 457-501.

Halle, M. 1995. To the editor. *Glot International* 1, 9/10.

Hayes, B. 1986/1987. A Revised Parametric Metrical Theory. *NELS* 17 (1986 Meeting), 274-289.

Hayes, B. 1995. *Metrical Stress Theory: Principles and Case Studies.* University of Chicago Press: London.

Howard, I. 1975. Can the 'Elsewhere Condition' Get Anywhere? *Language* 51, 109-127.

Janda, R. and M. Sandoval. 1984. "Elsewhere" in Morphology. IULC: Bloomington, Indiana.

Kenstowicz, M. 1994. *Phonology in Generative Grammar.* Blackwell: Cambridge.

Kiparsky, P. 1973. "Elsewhere" in Phonology. In *A Festschrift for Morris Halle*, S. Anderson and P. Kiparsky, eds., pp. 93-106.

Kiparsky, P. 1982. Lexical Phonology and Morphology. In *Linguistics in the Morning Calm*, The Linguistics Society of Korea, ed., Hanshin Publishing Co.: Seoul, pp. 3-91.

Kisseberth, C. 1970. The Functional Unity of Phonological Rules. *Linguistic Inquiry* 1.3, 291-306.

Koutsoudas, A., G. Sanders, and C. Noll. 1974. The application of phonological rules. *Language* 50, 1-28. (Drafted September, 1971: fn. 2).

McCarthy, J. & A. Prince. 1986. *Prosodic Morphology*. Ms., Brandeis University and UMass, Amherst. Reprinted, with corrections, as *Prosodic Morphology 1986*, TR-32, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick.

McCarthy, J. & A. Prince. 1995. Faithfulness and Reduplicative Identity. ROA-60, Rutgers Optimality Archive, http://ruccs.rutgers.edu/roa.html Condensed and extended as McCarthy & Prince 1997.

McCarthy, J. & A. Prince. 1997. Faithfulness and Identity in Prosodic Morphology. ROA-216, Rutgers Optimality Archive, http://ruccs.rutgers.edu/roa.html To appear in *The Morphology-Prosody Interface*, René Kager, Wim Zonneveld, and Harry van der Hulst, eds. CUP: London, Jerusalem, Athens.

Mester, A. 1995. The Quantitative Trochee in Latin. *Natural Language & Linguistic Theory* 12.1, 1-61.

Myers, S. 1987. Vowel Shortening in English. *Natural Language & Linguistic Theory* 5, 485-518.

Prince, A. 1985. Improving Tree Theory. *Proceedings of the Berkeley Linguistics Society* 11, 471-490.

Prince, A. 1990. Quantitative Consequences of Rhythmic Organization. In *CLS 26-II, Papers from the Parasession on the Syllable in Phonetics and Phonology*, K. Deaton, M. Noske, and M. Ziolkowski, eds. Chicago Linguistics Society: Chicago, pp. 355-398.

Prince, A. & P. Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. RuCCS-TR-2, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick. To appear, MIT Press: Cambridge.

Pullum, G. 1976. The Duke of York Gambit. *Journal of Linguistics* 12, 83-102.

Sanders, G. 1970. Precedence relations in language. Paper presented at summer meeting of the LSA, July 24.

Sanders, G. 1974. Precedence relations in language. *Foundations of Language* 11, 361-400.

Department of Linguistics
Rutgers University
18 Seminary Place
New Brunswick, NJ 08903