The acquisition of phonological opacity

Ricardo Bermúdez-Otero

University of Newcastle upon Tyne

Abstract. This paper argues that Stratal OT is explanatorily superior to alternative OT treatments of phonological opacity (notably, Sympathy Theory). It shows that Stratal OT supports a learning model that accounts for the acquisition of opaque grammars with a minimum of machinery. The model is illustrated with a case study of the classic counterbleeding interaction between Diphthong Raising and Flapping in Canadian English.

1. Phonological opacity: Stratal OT vs Sympathy Theory

Following the appearance of Prince & Smolensky (1993), phonologists were quick to realize that, in its original version, Optimality Theory (OT) was unable to describe a large set of phonological phenomena previously modelled by means of opaque rules. Ten years later, opacity remains the severest challenge confronting OT phonology. The problem is crucial because opacity effects constitute one of the clearest instances of Plato's Problem (Chomsky 1986) in phonology: learners face the task of acquiring generalizations that are not true on the surface. The ability to explain the acquisition of opaque grammars should accordingly be regarded as one of the main criteria by which generative theories of phonology are to be judged.

Among the variants of OT phonology currently on offer, two claim to provide a comprehensive solution to the problem of opacity: Sympathy Theory (McCarthy 1999, 2003) and Stratal OT (Bermúdez-Otero 1999b; Kiparsky 2000, forthcoming). In this paper, I compare the strategies whereby these two phonological models seek to achieve —and indeed transcend— explanatory adequacy:

- I shall first consider the formal restrictions that each theory imposes on the complexity of opaque effects (§2). In this area, several key principles of Sympathy Theory turn out to be conceptually problematic and/or empirically untenable.
- The main body of the article (§3-§7) focuses on the acquisition of opaque grammars. It will be shown that, whereas current approaches to acquisition in Sympathy Theory remain rudimentary (§7), Stratal OT supports an efficient learning algorithm that explains the acquisition of opacity effects with very little stipulation (§4). The effectiveness of this algorithm is illustrated with an application to the notorious contrast between [rəɪɾɪŋ] *writing* and [raɪɾɪŋ] *riding* in Canadian English (§5).
- Finally, I will suggest that, in its approach to non-paradigmatic opacity, Stratal OT transcends explanatory adequacy (Chomsky 2001), for the theory deals with non-paradigmatic opacity using mechanisms independently required by paradigm effects (§8).

2. Weak explanatory adequacy: typological restrictiveness

A theory of grammar is said to attain 'explanatory adequacy' when it solves the logical problem of language acquisition (Chomsky 1964, 1965). All too often, however, the term is used in a watered-down sense equivalent to 'typological restrictiveness': on the common assumption that learnability improves in proportion with reductions in the size of the grammar space generated by Universal Grammar (UG), grammatical frameworks that are typologically restrictive are often felt to be more explanatory (but cf. §3 below). In this section, therefore, I look at the space of possible opacity effects defined by Stratal OT (§2.1) and by Sympathy Theory (§2.2).

2.1. Stratal OT

Stratal OT borrows two key ideas from previous generative theories of phonology:

(1) **Cyclic application**
Given a linguistic expression e with a phonological input representation I, the phonological function P applies recursively from the inside out within a nested hierarchy of phonological domains associated with (but not necessarily fully isomorphic with) the morphosyntactic constituent structure of e:
i.e. if $I = [[x][[y]z]]$, then $P(I) = P(P(x),P(P(y),z))$.

(2) **Level segregation**
The phonology of a language does not consist of a single function P, but of a set of distinct functions or 'cophonologies' $\{P_1, P_2, \dots, P_n\}$, such that the specific function P_i applying to domains of type δ_i is determined by the type of morphosyntactic construction associated with δ_i (e.g. a stem, word, or phrase).

In this framework, opacity arises from the serial interaction between cycles. Within each cycle, however, the input-output mapping is transparent, as it is effected in the parallel fashion that characterizes classical OT:

(3) **Cycle-internal transparency**
Each cycle involves a single pass through *Gen* and *Eval*:
i.e. $P_i(\delta_i) = Eval_i(Gen(\delta_i))$

Principle (3) imposes severe restrictions on the complexity of opaque interactions. Notably, the depth of derivations is bound by the number of cycles, which is in turn independently constrained by the morphosyntactic structure of the linguistic expression. In addition, the phonology of the most inclusive domain (corresponding to processes applying across the board at the level of the Phonological Utterance) is predicted to be transparent.

Unlike rule-based Lexical Phonology and Morphology, however, Stratal OT does not invoke *ad hoc* principles to constrain the application of phonological processes. The principle of Structure Preservation, for example, plays absolutely no rôle in Stratal OT (other than as an empirically vacuous corollary of the optimization of underlying representations relative to the output of the highest stratum); see Bermúdez-Otero (1999b: 124). This results in major descriptive and explanatory gains. In Present-day English, for example, several phonological processes applying to stem domains have

traditionally been misascribed to the word level because of Structure Preservation. As shown by Bermúdez-Otero & McMahon (in preparation), rejecting Structure Preservation enables one to rectify these errors and re-establish a one-to-one correspondence between levels and domain types. Similarly, there is no room in Stratal OT for the Strict Cycle Condition, which is known to be empirically false (Kiparsky 1993).

2.2.   Sympathy Theory

In Sympathy Theory, apparent misapplication is caused by a set of constraints, called 'sympathy constraints', which enforce identity between the output and a failed co-candidate endowed with special status: the 'sympathetic candidate' (or '⊛-candidate'). This candidate is defined as the most harmonic among the subset of candidates satisfying a designated 'selector constraint' (or '✯-constraint').

   The theory, however, requires a number of additional stipulations. Unlike input-output faithfulness, for example, sympathy must be an asymmetric relationship: the output can copy properties of the ⊛-candidate but not vice versa, for otherwise opaque underapplication would be impossible (Bermúdez-Otero 1999b: 143-148). In contrast, IO-correspondence is symmetrical and reversible: outputs are faithful to the corresponding inputs in production, whereas in acquisition inputs are modelled upon outputs by Input Optimization (see §4.4 below). McCarthy (1999: 339) secures the asymmetry of sympathetic correspondence by means of the following stipulation:

> (4)   **Invisibility of sympathy constraints**
> Selection of sympathetic candidates is done without reference to sympathy constraints.

   Interestingly, this proviso imposes a significant restriction upon opacity effects. When two or more sympathetic candidates are active in a single computation, each is selected independently and affects the evaluation of output candidates in parallel. Sympathy Theory can therefore mimic serial derivations that involve at most one intermediate step:

$$(5) \quad \begin{array}{lcl} I \rightarrow a \rightarrow O & \approx & I \rightarrow ⊛ \rightarrow O \\ I \rightarrow a \rightarrow b \rightarrow O & \approx & ✗ \end{array}$$

Significantly, this empirical prediction turns out to be false: Bermúdez-Otero (2001, 2002b) adduces a counterexample from Catalan where two intermediate representations are crucially needed.

   Further to constrain the generative power of sympathy, McCarthy (1999: 339) adds another principle to the theory:

> (6)   ✯-*confinement*
> The selection of a sympathetic candidate must be confined to a subset of candidates that obey an IO-faithfulness constraint F.

This stipulation reduces the number of possible selector constraints and, therefore, the number of possible sympathetic relationships. In addition, it enables McCarthy to rationalize sympathy as a kind of 'faithfulness by proxy', where the optimal output copies some property of a hyperfaithful failed co-candidate.

   Empirically, however, the principle of ✯-confinement has been shown to cause undergeneration (Itô & Mester 1997; de Lacy 1998; Bermúdez-Otero 1999a, 1999b: 150-

191). The characterization of sympathy as 'faithfulness by proxy', moreover, does not translate into functional gains in terms of improved lexical access, for, as McCarthy (1999: 343) himself acknowledges, opaque processes are often neutralizing (see Bermúdez-Otero 1999b: 152).

In a final bid to restrict the complexity of sympathetic effects, McCarthy has also adopted special measures against non-paradigmatic non-vacuous Duke-of-York gambits. In serial terms, a Duke-of-York derivation has the form $a\rightarrow(\dots)\rightarrow b\rightarrow(\dots)\rightarrow a$; it is non-vacuous if b either escapes a process applicable to a ('bleeding') or undergoes a process not applicable to a ('feeding'); it is non-paradigmatic if b does not surface as (part of) a grammatically related expression (see §8 below). McCarthy claims that such derivations do not occur in natural language. To prevent Sympathy Theory from mimicking them, he resorts to a combination of two devices: one is the ✫-confinement clause stated in (6); the other is an *ad hoc* principle of 'cumulativity' (McCarthy 1999: §4.2; 2003), which penalizes output candidates that are more faithful to the input than the ✾-candidate.

Cumulativity is deeply problematic. First, it is simply false that non-paradigmatic non-vacuous Duke-of-York gambits do not occur in natural language. As shown in Bermúdez-Otero (2001, 2002b), such derivations do exist, and they are not hard to acquire provided that the phonological processes involved produce robust alternations (see §4-§6 below); one such case is found in Catalan. Secondly, the formal stipulations to which McCarthy resorts are fraught with difficulties. As we have seen, ✫-confinement is empirically untenable. In addition, it is only by brute force that the principle of cumulativity manages to block nonvacuous Duke-of-York gambits. Conceptually, moreover, cumulativity conflicts with the rationalization of sympathy as faithfulness by proxy.

3. Strong explanatory adequacy: the logical problem of language acquisition

As we have seen, Sympathy Theory fails in its attempts to define a highly restricted space of possible opacity effects. However, even if the theory attained this goal, the fact would be far less significant than McCarthy implies. This is because, in practice, typological restrictiveness does not guarantee explanatory adequacy in the strong sense. To appreciate this point, consider two theories of grammar T_1 and T_2, which define the grammar spaces S_1 and S_2 respectively. If both S_1 and S_2 are too large for convergence to be guaranteed by brute-force searching, then the prime determinant of learnability will be the relative efficiency of the learning algorithms associated with T_1 and T_2, rather than the relative size of S_1 and S_2 (see Tesar & Smolensky 2000: 2-3). In other words, a phonological model cannot achieve explanatory adequacy in respect of opacity simply by restricting the space of possible opaque effects; one must show that the learner is able to search that space effectively. Tellingly, there is to date no theory of the acquisition of sympathy-theoretic grammars (§7). In contrast, Stratal OT offers a straightforward recipe for the acquisition of opacity effects (§4-§6).

4. Phonological acquisition in Stratal OT: overview

This section presents the key ingredients for a model of phonological acquisition in Stratal OT. As the example in §5 will show, this model effectively accounts for the acquisition of opacity effects supported by evidence from alternations. The model achieves this by making the most of the assets of the synchronic theory: in particular, it fully exploits the serial interaction between strata and the intimate connection between the morphosyntactic domain of a phonological process and its stratal ascription (§4.1, §4.2). Beyond this, the model simply adopts current solutions to the problem of acquiring constraint rankings and input representations (§4.3, §4.4): the only provision added

specifically to deal with opaque phenomena is the principle of Archiphonemic Prudence introduced in §4.5.

4.1. Iterative stratum construction

Stratal OT enables one to break the logical problem of phonological acquisition down into a set of relatively simpler subproblems, for learning a phonological grammar consists of acquiring a series of cophonologies: typically, the phrase-level, word-level, and stem-level cophonologies. Moreover, since the input to level n provides the output of level $n-1$, each of these subproblems can be tackled in a logical progression. Acquiring the phrase-level cophonology, for example, involves (i) discovering the phrase-level constraint hierarchy and (ii) assigning single representations to individual words at the input to the phrase level. The input representations so assigned constitute the output of the word level and provide the data for the next iteration in the process of acquisition. Thus, as Bermúdez-Otero (1999b: 102) puts it, "the learner of an interleaved grammar does not tackle all alternations *en masse*, but rather peels away levels of morphosyntax one by one like onion layers".

4.2. The emergence of opacity

As we saw in §2.1, opacity arises from interactions between processes that apply transparently in their own strata: each phonological generalization in the grammar holds true in the output of the corresponding level, which defines the domain of the generalization. During acquisition, therefore, the task of assigning phonological processes to the appropriate strata can be reduced to the independent problem of discovering correct input representations.[1] Consider, for example, a process p that applies at level n and is rendered opaque by changes introduced at level $n+1$. If input representations are correctly assigned at level $n+1$, p will be true of the output of n. On this basis, any of the standard ranking algorithms designed to acquire transparent processes will establish the ranking for p in the constraint hierarchy of n. By the same token, the constraint ranking for p will not be introduced at level $n+1$ simply because the ranking algorithm encounters contradictory data, as p does not hold true in the output of $n+1$. In other words, the learning model need do no more than establish transparent constraint rankings (§4.3) and assign input representations correctly (§4.4, §4.5); the grammatical architecture of Stratal OT takes care of the rest.

4.3. Constraint ranking by pure phonotactic learning under the identity map

At any level, then, the first task for the learner is to find the appropriate ranking of constraints, given a set of output forms. As Prince & Tesar (1999) and Hayes (1999) have shown, this can be done largely on the basis of purely distributional information: assuming the identity map (input = output) plus a MARKEDNESS » FAITHFULNESS bias (henceforth, 'M » F bias'), the learner must demote markedness constraints and promote faithfulness constraints just enough to derive the output from identical input. The details of the ranking algorithm need not concern us here. The important point, rather, is that alternations usually conspire to bring morphological or syntactic collocations in line with output phonotactics (Kisseberth 1970); for this reason, pure phonotactic learning will in most cases suffice to find the constraint rankings driving not only phonotactics but also

---

[1] In Stratal OT, only the highest grammatical level is subject to Richness of the Base. The input to a non-initial stratum n will possess systematic properties enforced by the constraint hierarchy of level $n-1$ (Bermúdez-Otero 2002b: §24-§26; Bermúdez-Otero & Hogg forthcoming: note 16).

alternations. The acquisition of the latter will then boil down to mere input assignment (Hayes 1999: §6).[2]

4.4. *Input assignment (I): alternations prompt departures from the identity map*

Following the currently prevalent view, I assume that learners need evidence from alternations in order to depart from the identity map (see e.g. Yip 1996 and §6 below). In line with the principle of Input Optimization (Prince & Smolensky 1993: §9.3),[3] departures from the identity map are minimal: unwarranted disparity between inputs and outputs causes unnecessary violations of faithfulness constraints. Unfortunately, currently available formulations of Input Optimization for alternating items (e.g. Inkelas 1995) are flawed; see Inkelas (2000: §5.2) for discussion of some of the problems. Bermúdez-Otero (in preparation) develops a fully-fledged alternative, supported with diachronic evidence from changes involving input restructuring (see §6).

We cannot go into details here, but, in brief, Bermúdez-Otero (in preparation) characterizes optimal inputs as follows:

(7)   **Input optimality (after Bermúdez-Otero in preparation)**
An input representation is optimal iff it has no competitor that
  • generates an identical set of output alternants,
  • generates all output alternants no less efficiently,
and   • generates some output alternant more efficiently.[4]

In practice, this definition of input optimality selects a set J of potential inputs whose members are all output-equivalent and where each member is maximally similar to some output alternant. If the cardinality of J is greater than 1, the learner can make a (provisional) choice among its members by means of certain heuristics:

(8)   **Hale's heuristic (after Hale 1973: 420)**
Prefer inputs that are well-formed outputs.

(9)   **Heuristic for asymmetric paradigms**
In an asymmetric paradigm, prefer those inputs which generate the central member of the paradigm most efficiently.

In (9), the term 'paradigm asymmetry' refers to the well-known observation that citation forms often enjoy a special status in comparison with sandhi forms, that the nominative singular may be more central than other members of nominal paradigms, and so forth (see e.g. Kuryłowicz 1949, Mańczak 1958, Lahiri 1982, Lahiri & Dresher 1983-84).

---

[2]   In some cases, constraints can be ranked appropriately only if the correct input representations are known. This problem can usually be solved by iterating between constraint ranking and input assignment until equilibrium is reached (see Tesar & Smolensky 2000: §1.3.2, §5.2); but see also §6 below.

[3]   The original term 'Lexicon Optimization' is inappropriate in Stratal OT, where unmotivated disparity between inputs and outputs is avoided in all strata but inputs coincide with underlying representations only at the highest level (Bermúdez-Otero & Hogg forthcoming: note 2).

[4]   More efficient inputs cause fewer violations of high-ranking faithfulness constraints. Note that input choice can only affect faithfulness, as markedness constraints only evaluate output forms.

4.5. Input assignment (II): Archiphonemic Prudence

The final task for the learner is to assign input representations to non-alternating items. At this point, it is essential for the acquisition of opaque grammars that the learner should be able to use evidence from alternations to detect deviations from the identity map in non-alternating items. I suggest that this can be achieved by supplementing current learning models with a principle of 'Archiphonemic Prudence', designed to deal with possible instances of neutralization in non-alternating environments.

Let there be two input elements /α/ and /β/ at level n, such that, in the output of n, the contrast between /α/ and /β/ is maintained in environment [___]$_e$ and neutralized in environment [___]$_f$. Let γ be the output realization of /α/ and /β/ in the neutralizing environment [___]$_f$. In such circumstances, the output of n will contain alternations such as [α]$_e$~[γ]$_f$ and [β]$_e$~[γ]$_f$. We may refer to any token of [γ]$_f$ in the output of n as an 'archiphonemic string'. The problem arises when the learner comes across such an archiphonemic string in a non-alternating item i.[5]

I propose that, under Archiphonemic Prudence, the learner relies on the evidence from alternations such as [α]$_e$~[γ]$_f$ and [β]$_e$~[γ]$_f$ to assign an input representation to i at level n. First, the learner creates two potential representations for i in the input to n: one where the input correspondent of γ is /α/, and another where the input correspondent of γ is /β/. The input candidates are otherwise identical with the output realization of i (recall that deviations from the identity map are minimal). These input candidates are then 'quarantined': they are not included in the data set triggering phonological acquisition at level $n-1$; learning at $n-1$ proceeds exclusively on the basis of non-quarantined inputs to n.[6] When the constraint hierarchy of level $n-1$ is known, the learner is in a position to choose between the two quarantined candidates for input representation of i at level n: if the input candidate containing /α/ is not a well-formed output at level $n-1$, the learner chooses the input candidate containing /β/.[7]

5. Case study: Diphthong Raising and Flapping in Canadian English

In this section, the learning model outlined in §4 is applied to a classic empirical problem from Canadian English: the opaque interaction whereby the Flapping of /t/ (which also applies to /d/) counterbleeds the Raising of /aɪ/ and /aʊ/ to [əi] and [ʌʊ] before voiceless obstruents. As is well-known, this counterbleeding effect results in the apparent overapplication of Raising on the surface:[8]

(10)		*writing*	*riding*	*mitre*	*powder*
	UR	/raɪt-ɪŋ/	/raɪd-ɪŋ/	/maɪtər/	/paʊdər/
	Raising	rəitɪŋ	—	məitər	—
	Flapping	**rəiɾɪŋ**	raɪɾɪŋ	**məiɾər**	paʊɾər

---

[5]   As we shall see in §5.2, the learner can identify archiphonemic strings by examining sets of output alternants and factoring out the portions shared by all the members of each set.

[6]   The set of non-quarantined inputs will consist of input representations for alternating items, as well as input representations for non-alternating items not containing archiphonemic strings.

[7]   If both candidates are possible outputs at level $n-1$, they remain quarantined and the choice is passed on to level $n-2$.

[8]   In transcriptions, I ignore all allophonic detail not directly relevant to the discussion. In my choice of symbols for the diphthongs, I follow Wells (1982: §6.2.4). I am deeply grateful to my colleague Dr John Stonham for acting as a native speaker informant and for discussing with me the analysis presented in §5.1.

Accounting for the acquisition of this opaque interaction is a highly significant result. Since it was first highlighted by Joos (1942), the problem has figured prominently in the theoretical debate: see Chomsky (1964: 74), Chomsky & Halle (1968: 342), and Bromberger & Halle (1989: 58-61), among others. Kenstowicz (1994: 6-7) discusses it as a canonical example of Plato's Problem in phonology and, significantly, Hayes (1999: §8) uses it to illustrate the challenges of learning morphophonological alternations in OT.

5.1.   The target grammar

In this section I describe the grammatical system that learners of Canadian English must acquire. For the sake of concreteness, I assume foot-based analyses for both Flapping and Diphthong Raising (Kiparsky 1979, Jensen 2000). This choice, however, is irrelevant to the application of the learning model, which would operate in exactly the same way under an analysis based on ambisyllabicity (Kahn 1976, Gussenhoven 1986).

Flapping involves the realization of /t/ and /d/ as [ɾ] when (i) lax, (ii) preceded by a vowel or [r], and (iii) followed by a vowel. I assume, following Jensen (2000), that /t/ and /d/ are tensed at the word level if foot-initial; otherwise, they are lax (and so extra-short). Crucially for our purposes, Flapping is phrase-level, as indicated by the fact that it applies when its environment straddles a word boundary, as in (11c) and (11d):[9]

(11)	a.	[fæɾər]	*fatter*	cf.	[fæt]	*fat*
	b.	[mæɾər]	*madder*	cf.	[mæd]	*mad*
	c.	[hi hɪɾ æn]	*he hit Ann*	cf.	[hɪt]	*hit*
	d.	[hi hɪɾ æn]	*he hid Ann*	cf.	[hɪd]	*hid*

In the sentence given in (11c), the /t/ of *hit* is lax because it is not foot-initial at the word level; the /t/ only becomes prevocalic (and, in this case, also foot-initial by resyllabification) at the phrase level, where the words in the sentence are concatenated.

The diphthongs /aɪ/ and /aʊ/ undergo Raising to [əi] and [ʌʊ] when followed by a voiceless obstruent in the same foot.[10] The examples in (12a) illustrate the rôle of consonant voicing; those in (12b), the rôle of foot structure.

(12)	a.	[nəif]	*knife*	cf.	[naɪvz]	*knives*
		[hʌʊs]	*house*	cf.	[haʊzɪz]	*houses*
	b.	[ˈsəifn]	*syphon*	cf.	[saɪˈfanɪk]	*syphonic*
		[səit]	*cite*	cf.	[saɪˈteɪʃn]	*citation*

I suspect that, historically, Raising arose through the phonologization of a qualitative side effect of 'Pre-Fortis Clipping' (the shortening of vowels before fortis obstruents). Informally, I assume that the constraint hierarchy that enforces Raising includes a context-free markedness constraint CLEARDIPH, which favours diphthongs where the auditory distance between the two elements is maximal; this constraint penalizes [əi] and [ʌʊ]. In the environment of Pre-Fortis Clipping, however, the context-sensitive markedness constraint CLIPDIPH demands that the distance between diphthongal

---

[9]   The stratal ascription of Flapping is well-known (see e.g. Kaisse & Shaw 1985: 4ff.). I substitute the term 'phrase-level', which neutrally designates the domain of the process, for the more loaded term 'postlexical'.

[10]   For our purposes, we could just as well assume an analysis where underlying /əi/ and /ʌʊ/ undergo lowering to [aɪ] and [aʊ] everywhere except before voiceless obstruents in the same foot; for our learning model, the choice is immaterial (see notes 1 and 17).

elements should be minimized, thereby penalizing [aɪ] and [aʊ].[11] To be active, CLEARDIPH must dominate its faithfulness antagonist IDENT[mid], whilst CLIPDIPH must dominate IDENT[low]:[12]

(13) **IDENT[mid]**
Let α be an input segment, and let β be its output correspondent;
if α is [mid], then β is [mid].

IDENT[low]
Let α be an input segment, and let β be its output correspondent;
if α is [low], then β is [low].

In addition, the context-sensitive markedness constraint must dominate its context-free counterpart. Thus, the normal application of Raising requires each of the following three rankings:
• CLEARDIPH » IDENT[mid],
• CLIPDIPH » IDENT[low],
and • CLIPDIPH » CLEARDIPH.

Crucially, there is clear evidence that Raising is 'lexical' (i.e. not phrase-level), as diphthongs are not raised when a voiceless obstruent follows across a word boundary:

(14)   ['laɪ fər mi]   *lie for me*
    cf.   ['ləifər]   *lifer* (i.e. 'convict serving a life sentence')

In fact, Raising probably applies at the stem level. First, word-level suffixes such as *-ful* and *-ship* do not trigger Raising:

(15) ['aɪfʊl]   *eyeful*[13] cf. ['əifəl] *Eiffel (Tower)*
      ['fraʊʃɪp], *['frʌʃɪp]   *Frauship* (nonce word derived from German *Frau* on the analogy of *lordship, ladyship*)

Secondly, Raising has lexical exceptions for some speakers (Wells 1982: 495, citing Chambers 1973). Such behaviour is most often observed among phonological processes applying at the highest level in the grammar:[14]

(16) ['saɪklɑps] *Cyclops* vs ['məikrɑn] *micron*

Finally, recall that Structure Preservation plays no rôle in Stratal OT (see §2.1 above) and cannot therefore be invoked as an argument against locating Raising in the stem level.

---

[11] For markedness constraints on diphthongs, see e.g. Minkova & Stockwell (forthcoming).

[12] IDENT[mid] and IDENT[low] presuppose monovalent height features, but could be restated in terms of the plus and minus values of a binary feature [±low]; see Pater (1995). Feature-value faithfulness has been criticized in, for example, Baković (1999, 2000). In the case of Canadian English, an analysis based on a symmetrical constraint IDENT[±low] would require learners to follow a slightly different learning path to that described in §5.2-§5.4 below, but would not be an obstacle to successful convergence.

[13] In this example, Raising is unlikely to be blocked by a weak foot over *-ful*. The word seems to be metrically equivalent to the univerbated compound ['həiskul] *high school*, where Raising does apply (see Wells 1982: 494); cf. the unfused variant ['haɪ ˌskul] .

[14] Unexpected instances of constrastive raising in some idiolects, such as ['əiɾəl] *idle* vs ['aɪɾəl] *idol* (Hayes 1999: §8.5, citing Vance 1982), constitute another aspect of the same phenomenon.

In sum, Diphthong Raising applies to stem domains, whereas the domain of Flapping is phrasal. From this information, Stratal OT correctly derives their relative order of application: phrase-level Flapping must follow —and will therefore counterbleed— stem-level Raising.

How, then, can this grammatical system be acquired using the learning model described in §4?

- Setting up the constraint hierarchy for Flapping at the phrase level is clearly the easiest task: since Flapping is surface-true, the learner can achieve this by pure phonotactic learning from the primary data.

- In the case of Raising, in contrast, instances of surface overapplication (e.g. *writing, mitre*) and underapplication (e.g. *eyeful, lie for me*) will prevent the learner from establishing a raising hierarchy at the phrase level.

- Next, the learner must use the evidence from phrasal alternations such as *hit* vs *hit Ann* and *hid* vs *hid Ann* to discover the fact that surface [ɾ] derives from either /t/ or /d/ in the output of the word level, but —crucially— not from */ɾ/.

- In addition, the learner must be able to capitalize on this information and, using Archiphonemic Prudence, avoid the incorrect identity map */ɾ/→[ɾ] in non-alternating items such as /maɪtər/→[məiɾər] *mitre* and /vaɪtəl/→[vəiɾəl] *vital*.

- If the learner chooses the correct input representations for alternating items at the phrase and word levels, Raising will become output-true at the stem level, and the learner will be able to establish the constraint ranking for Raising in the stem-level hierarchy by pure phonotactic learning.

- At this point, the learner can turn to items such as *mitre* and *vital*, previously quarantined under Archiphonemic Prudence. Since the stem-level constraint hierarchy enforces normal application of Raising, the incorrect phrase- and word-level inputs */məidər/ and */vəidəl/ can be discarded, as they are ill-formed stem-level outputs. This just leaves the target input representations with /t/.

The success of this account rests upon two simple ideas. First, the constraint ranking driving a process *p* is established in the hierarchy of level *n* if and when *p* is true in the output of *n*; thus, the contrast between normal application and misapplication enables learners to assign phonological processes to the correct strata (§4.2). Secondly, learners depend on alternations to depart from the identity map either directly (in the case of alternating items; §4.4) or indirectly (when required by Archiphonemic Prudence; §4.5).

5.2.   Acquiring the phrase-level cophonology

If we ignore the problem of covert structure (see e.g. Tesar & Smolensky 2000: 6ff.), the primary linguistic data provide the child with direct access to the phrase-level output. Applying pure phonotactic learning to these data, the child will be able to establish the ranking for Flapping in the phrase-level constraint hierarchy, as Flapping is surface-true. In contrast, table (17) shows how the surface misapplication of Diphthong Raising prevents the leaner from establishing the rankings CLEARDIPH » IDENT[mid] and CLIPDIPH » CLEARDIPH, which, as we saw in §5.1, are essential to the process.

(17)

Datum	Triggered ranking
məiɾər › maiɾər 'mitre' rəiɾiŋ › raiɾiŋ 'writing'	IDENT[mid] » CLEARDIPH
aiful › əiful 'eyeful' lai fər mi › ləi fər mi 'lie for me'	CLEARDIPH » CLIPDIPH[15]

Next, the child must undo phrase-level alternations by assigning a single representation to each word in the phrase-level input. Note that, at this stage, the learner does not yet attempt to analyse word-level collocations such as *writ-ing*, *rid-ing*, and *eyeful*; at the phrase level, these are treated in the same way as monomorphemic items like *mitre* and *powder*.

Let us first consider the alternation [hɪt] *hit* ~ [hɪɾ æn] *hit Ann*. If we assume minimum disparity between inputs and outputs (§4.4), there is only one possible phrase-level input representation for *hit*: viz. /hɪt/. Note that */hɪɾ/ and */hɪd/ would both incorrectly generate [hɪd]~[hɪɾ æn], as the phrase-level constraint hierarchy does not neutralize voice contrasts in word-final position. Crucially, by factoring out the identical portion of the alternants [hɪt]~[hɪɾ], the learner discovers a set of alternating elements [t]~[ɾ]. And, given /hɪt æn/→[hɪɾ æn], she finds out that /t/ is a possible phrase-level input representation for [ɾ] in the flapping environment.

Let us now turn to [hɪd] *hid* ~ [hɪɾ æn] *hid Ann*. Here, the set *J* of optimal phrase-level inputs for *hid* consists of two members: viz. /hɪd/ and */hɪɾ/ (§4.4). Since [d] and [ɾ] are in complementary distribution on the surface, both representations generate the correct set of output alternants. In this case, however, both Hale's heuristic (8) and the heuristic for asymmetric paradigms (9) favour input /hɪd/. Since the learner has no reason to retract this hypothesis, */hɪɾ/ is discarded. On this basis, the child discovers a new alternating set [d]~[ɾ] derived from input /d/.

The child now knows that [ɾ] in the Flapping environment is an archiphonemic string with two possible input correspondents: /t/ or /d/; see §4.5 above. By Archiphonemic Prudence, therefore, non-alternating items such as *mitre*, *powder*, *writing*, and *riding* must be quarantined, and the assignment of phrase-level input representations to them is deferred. Assuming that the learner countenances the minimal departure from the identity map compatible with Archiphonemic Prudence, the choice of inputs will be as in (18):

(18)

Quarantined item	Phrase-level input candidates
[məiɾər] 'mitre'	/məitər/, /məidər/
[paʊɾər] 'powder'	/paʊtər/, /paʊdər/
[rəiɾiŋ] 'writing'	/rəitiŋ/, /rəidiŋ/
[raiɾiŋ] 'riding'	/raitiŋ/, /raidiŋ/

5.3. Acquiring the word-level cophonology

Leaving aside the quarantined items in (18), the child can now proceed to the acquisition of the word-level cophonology. At this stage, the data set consists of the single whole words that remain in the non-quarantined phrase-level input: e.g. /hɪt/ *hit*, /hɪd/ *hid,* /rəit/

---

[15] Under M » F bias (see §4.3), it is preferable to impute violations of CLIPDIPH to a higher-ranked markedness constraint (i.e. CLEARDIPH), rather than to faithfulness (i.e. IDENT[low]).

write, /raɪd/ *ride*, /aɪfʊl/ *eyeful*, etc. Crucially, there is no form in this data set where either [əi] or [ʌʊ] fails to be followed by a voiceless obstruent in the same foot. Recall that all items in which Raising overapplies word-internally, such as [məiɾər] *mitre* and [rəiɾɪŋ] *writing*, have been placed under quarantine. On the surface, Raising also overapplies in forms subject to Flapping across word boundaries: e.g. [rəiɾ ʌp] *write up*. These forms, however, are involved in phrase-level alternations (e.g. [rəit] *write* ~ [rəiɾ ʌp] *write up*) and consequently disappear in the processes of phrase-level input assignment. Remember that, at phrase level, [rəiɾ ʌp]←/rəit ʌp/ (see §5.2 again). Nonetheless, even if Raising no longer overapplies, there still remain instances of underapplication: e.g. [aɪfʊl] *eyeful*.

Let us now consider the outcome of pure phonotactic learning in this situation. Since the data include raised diphthongs, CLEARDIPH must be crucially dominated, either by CLIPDIPH or by IDENT[mid]. Note, however, that all violations of CLEARDIPH occur before voiceless obstruents in the same foot, for, as we have just seen, there is no overapplication of Raising in the non-quarantined data. Accordingly, a learner subject to M » F bias will respond to the datum rəit › raɪt by ranking CLIPDIPH above CLEARDIPH, whilst preserving the default ranking CLEARDIPH » IDENT[mid]. In contrast, the datum aɪfʊl › əifʊl cannot be imputed to a contextual markedness effect, and so triggers the ranking IDENT[low] » CLIPDIPH. This results in the hierarchy IDENT[low] » CLIPDIPH » CLEARDIPH » IDENT[mid].[16]

At this point, the quarantine on nonalternating items such as *mitre* and *writing* may be lifted, as the newly established word-level hierarchy forces a choice between the phrase-level input candidates allowed by Archiphonemic Prudence. Observe that *[məidər] and *[rəidɪŋ] are ill-formed word-level outputs because they show overapplication of Raising. These forms cannot therefore be derived from identical input under the word-level ranking IDENT[low] » CLIPDIPH » CLEARDIPH » IDENT[mid]. In consequence, the phrase-level input representations for *mitre* and *writing* must be /məitər/ and /rəitɪŋ/, respectively; see (18) above.

(19)

		IDENT[low]	CLIPDIPH	CLEARDIPH	IDENT[mid]
məidər	məidər			*!	
	maɪdər ☞				*
məitər ☞	məitər ☞			*	
	maɪtər		*!		*
rəidɪŋ	rəidɪŋ			*!	
	raɪdɪŋ ☞				*
rəitɪŋ ☞	rəitɪŋ ☞			*	
	raɪtɪŋ		*!		*

---

[16] The data are also compatible with less restrictive rankings such as IDENT[mid] » CLEARDIPH » CLIPDIPH, which not only permits unraised diphthongs in the environment of Raising (e.g. [aɪfʊl] *eyeful*), but also tolerates raised diphthongs in *any* environment and not just before voiceless obstruents in the same foot: e.g. *[rəid]. I assume, however, that the constraint ranking algorithm always selects the most restrictive hierarchy. This is not the place to discuss whether or not current implementations of the M » F bias reach this goal in every circumstance. Suffice it to say that an effective account of the acquisition of opaque grammars presupposes a successful solution to the Subset Problem in pure phonotactic learning; see Prince & Tesar (1999) for much germane discussion.

In contrast, it is not yet possible at this stage to lift the quarantine on *powder* and *riding*. In this case, the incorrect phrase-level inputs are */paʊtər/ and */raɪtɪŋ/, which contain unraised diphthongs followed by a voiceless obstruent in the same foot. However, since underapplication of Raising is still tolerated at the word level (cf. [aɪfʊl] *eyeful*), both these forms are possible word-level outputs —as are the targets /paʊdər/ and /raɪdɪŋ/. The choice of input for *powder* and *riding* must accordingly wait until the stem-level constraint hierarchy is known (see note 7).

Nonetheless, the lifting of the quarantine on *mitre* and *writing* frees up more data for pure phonotactic learning at the word level. The new word-level output forms, e.g. [məitər] *mitre* and [rəitɪŋ] *writing*, are counterexamples to Flapping and therefore enable the child to learn that Flapping does not apply at the word level (or higher in the grammar). Inexorably, more and more bits of grammatical knowledge fall into place.

The child can now turn to input assignment. This is pretty straightforward at the word level, as the partial lifting of the quarantine has not revealed new alternations. Accordingly, the learner has no reason to deviate from the identity map: i.e. /hɪt/→[hɪt], /hɪd/→[hɪd], /məitər/→[məitər], etc. In particular, word-level derivatives such as [aɪfʊl] *eyeful* and [rəit-ɪŋ] *writing* do not create alternations with their respective base forms: cf. [aɪ] *eye* and [rəit] *write*. The input representation of the stem will therefore be identical with its output realization: i.e. /aɪ-/ *eye* and /rəit-/ *write*.

5.4.   Acquiring the stem-level cophonology

By this time, the learner has taken a decisive step forward: in effect, when she removes word-level suffixes such as *-ful* and *-ship* from collocations such as [aɪfʊl] *eyeful* and [fraʊ-ʃɪp] *Frauship* (see (15) above), she disposes of the last remaining instances of Raising misapplication. As we saw in §5.3, the input to the word level consists of:
  • monomorphemic items such as /ˈməitər/ *mitre*, /rəit/ *write*, /səit/ *cite*, /ˈsəifn/ *syphon*, /aɪ/ *eye*;
and • stem-level collocations such as the irregular verbs /hɪt/ *hit* and /hɪd/ *hid*, or the level-one derivatives /saɪˈfɑnɪk/ *syphonic* and /saɪˈteɪʃn/ *citation*.
These forms, which provide the trigger for phonological acquisition at the stem level, obey Raising. In consequence, Raising becomes true of the stem-level output, and the appropriate constraint ranking can be installed in the stem-level hierarchy by pure phonotactic learning.

At last, the child can lift the quarantine on *powder* and *riding*. The newly acquired stem-level hierarchy successfully discards the incorrect phrase-level inputs */paʊtər/ and */raɪtɪŋ/, where Raising underapplies. In consequence, /paʊdər/ and /raɪdɪŋ/ are returned as the phrase-level input representations for *powder* and *riding*.

(20)

		CLIPDIPH	CLEARDIPH	IDENT[low]	IDENT[mid]
paʊtər	paʊtər	*!			
	pʌʊtər ☞		*	*	
paʊdər ☞	paʊdər ☞				
	pʌʊdər		*!	*	
raɪtɪŋ	raɪtɪŋ	*!			
	rəitɪŋ ☞		*	*	
raɪdɪŋ ☞	raɪdɪŋ ☞				
	rəidɪŋ		*!	*	

At the word level, the child can now sort out the paradigm [raɪd] *ride* ~ [raɪd-ɪŋ] *riding*. Since the paradigm proves to be non-alternating, the child adheres to the identity map and selects /raɪd-/ as the input representation of the stem. There then remains the task of identifying the input to the stem level, but no special difficulty arises here.[17] For all intents and purposes, the acquisition of the counterbleeding interaction between Diphthong Raising and Flapping in Canadian English is now complete.

6. Stratal OT and analogical change

As I have repeatedly stated, it is essential for the success of the learning algorithm that the child should depart from the identity map only when presented with evidence from alternations (§4.4-§4.5, §5.1). This is a simple corollary of Input Optimization, which, in turn, results from applying to input assignment the fundamental OT idea that constraint violation is minimal. Significantly, if the child adheres to the identity map unless confronted with alternations, it is predicted that the obliteration of alternations by phonological change in the output of level *n* will cause restructuring in the input to *n*. As shown in Bermúdez-Otero (2002a, in preparation) and Bermúdez-Otero & Hogg (forthcoming), this straightforward mechanism explains most instances of analogical change.

One should bear this diachronic result in mind when evaluating the effectiveness of the learning algorithm presented in §4. Notably, the procedure for assigning input representations to alternating items (§4.4) may appear particularly prone to error. This apparent weakness of the theory, however, proves to be one of its greatest strengths, as errors in input assignment (leading to input restructuring) are the main engine of morphophonological change; see Hayes (1999: §8.5) for similar suggestions.

In this connection, I feel sceptical about Alderete & Tesar's (2002) claim that deviations from the identity map may be prompted by purely distributional evidence. The alleged instances of this phenomenon —which Alderete & Tesar mention but do not discuss— should be scrutinized with the utmost rigour to determine whether they are (i) psychologically real and (ii) diachronically stable. Only if the empirical evidence is solid and abundant should we seek to endow the learner with more power to adopt abstract inputs. If we move too far in this direction, we shall find ourselves in the dilemma, familiar from early rule-based theory, of lacking a principled account for changes in underlying representations.[18]

7. The acquisition of opacity in Sympathy Theory

Stratal OT supports a learning algorithm that explains the acquisition of opaque grammars with minimal stipulation and successfully accounts for the incidence of analogical change. In stark contrast, there is at present no algorithm for acquiring sympathy-theoretic grammars (McCarthy 1999: 340). Dinnsen *et al.* (2000) use Sympathy Theory to analyse certain opacity effects in the speech of children suffering from phonological delay, but their contribution is more descriptive than explanatory.

---

[17] This job involves undoing transparent alternations such as [səɪt] *cite* ~ [saɪˈteɪʃn] *citation* under Richness of the Base, just as in classical monostratal OT. See notes 1 and 10.

[18] Bermúdez-Otero (in preparation) uses a similar argument to rebut Halle & Idsardi's (2000) charge that OT allows constraint rankings that are too delicate and specific and, therefore, unlearnable. In an examination of morphophonological change in *a*-stem nouns in Old English, Bermúdez-Otero (in preparation) shows that it was precisely the delicate and specific rankings that became subject to loss through analogical change.

Indeed, their suggestions regarding the acquisition of sympathy-theoretic grammars are sketchy and hold little promise of success:

- First, Dinnsen *et al.* (2000: §2.7) propose an inductive bias in favour of ranking sympathy constraints above IO-faithfulness. This brute force device is needed to compensate for the fact that, in Sympathy Theory, there is no obvious trigger for the acquisition of opaque grammars, as the theory severs the essential link between alternations and non-paradigmatic opacity (§6, §8). From a diachronic viewpoint, Dinnsen *et al.*'s suggestion is likely to befog our understanding of analogical change. The evidence shows that, historically, surface properties derived opaquely (i.e., in Dinnsen *et al.*'s terms, by high-ranking sympathy constraints) tend to be reanalysed as underlying (i.e. as generated, in Dinnsen *et al.*'s terms, by high-ranking IO-faithfulness). Dinnsen *et al.*'s proposals predict precisely the opposite.

- Secondly, Sympathy Theory raises a serious learnability question: since every crucially active IO-faithfulness constraint can be a selector constraint, the learner has to contend with a large bouquet of potential ✿-candidates for each output form. To deal with this 'bouquet problem', Dinnsen *et al.* (2000: §3.1) appeal to the Elsewhere Condition: they suggest that, when a potential ✿-candidate constitutes a superset of another (in representational terms), the more specific (or superset) representation should be chosen as the sympathetic candidate. The rationale for this suggestion is obscure, other than that it happens to work for the particular example that Dinnsen *et al.* discuss.

In this light, Sympathy Theory offers little hope of accounting for the acquisition of opaque grammars: it fails to attain either weak or strong explanatory adequacy (§2, §3).

8. Beyond explanatory adequacy

As we have seen, Sympathy Theory's most serious flaw lies in its failure to capture the vital link between opacity in non-alternating items and paradigmatic misapplication: the former is handled by sympathy constraints, whereas the latter falls to OO-correspondence (Benua 1997; see McCarthy 2003). It is this flaw that leads Sympathy Theory to focus on mere typological restrictiveness (§2.2) and prevents it from aspiring to strong explanatory adequacy (§7).

Stratal OT, in contrast, makes the most of the link between paradigmatic and non-paradigmatic opacity. From a descriptive viewpoint, opacity in non-alternating items is handled through cyclic application and level segregation; see (1) and (2). Both devices are independently required to deal with the misapplication effects (typically associated with alternations) that arise from morphological domain restrictions on phonological processes. As regards learnability, moreover, the same procedures that underpin the acquisition of alternations enable the child to learn opaque derivations in non-alternating environments (§4.1-§4.4); the only special stipulation required by the latter is the principle of Archiphonemic Prudence (§4.5). This intimate bond between paradigmatic and non-paradigmatic opacity is strikingly illustrated by the counterbleeding interaction analysed in §5: it is alternations such as [rəit] *write* ~ [rəiɾ ʌp] *write up* vs [raɪd] *ride* ~ [raɪɾ ʌp] *ride up* that enable the child to discover the correct underlying representations for non-alternating items such as [məiɾər] *mitre* and [pɑʊɾər] *powder*.

This contrast also determines the extent to which Sympathy Theory and Stratal OT may claim to transcend explanatory adequacy. As shown by Chomsky (2001), competing theories of grammar may be assessed at a level higher than that of explanatory adequacy. If one finds that a linguistic theory *T* solves the logical problem of language acquisition, one should then ask whether *T* can account for the contents of UG; insofar as *T* explains

why UG is the way it is, *T* can be said to transcend explanatory adequacy. This goal can be attained by means of two main strategies, which —despite linguists' visceral inclinations— are not mutually incompatible. The 'minimalist' strategy involves paring UG down to the absolute minimum: the less UG contains, the less there is to explain. The 'evolutionary' strategy involves providing an account of the phylogenesis of the language faculty that explains the composition of UG. Typically, particular features of UG will be claimed to have arisen either through adaptation or through exaptation.

Sympathy Theory is plainly not a minimalist theory of opacity. As I have shown, it postulates different UG mechanisms to deal with paradigmatic and non-paradigmatic opacity. Accordingly, if Sympathy Theory ever attained explanatory adequacy, it could only hope to transcend that level through the evolutionary route. However, I cannot think of a plausible way to explain the phylogenesis of sympathy constraints. One might suggest that, insofar as sympathetic correspondence enhances lexical recognition, it confers an adaptive advantage upon speakers, which might have led to sympathy becoming hardwired during the evolution of the species. However, the evidence reviewed in §2.2 shows that the characterization of sympathetic effects as 'faithfulness by proxy' fails to support an adaptive scenario for the philogenesis of sympathy constraints: notably, 'faithfulness by proxy' relies on the principle of ✱-confinement, which is empirically untenable; it creates an unresolved tension with the notion of cumulativity; and it fails to provide a consistent functional advantage in lexical access.

Unlike Sympathy Theory, Stratal OT transcends explanatory adequacy by the minimalist route: it dispenses with *ad hoc* UG devices to handle non-paradigmatic opacity, reducing the latter to an epiphenomenon of the interface of phonology with morphology and syntax (*modulo* Archiphonemic Prudence). In sum, Stratal OT solves Plato's Problem in relation to opacity effects, and does so with the bare minimum of stipulation. Surely, this is the kind of phonological theory we want.

References

[ROA = Rutgers Optimality Archive, http://roa.rutgers.edu/]

Alderete, J. & B. Tesar (2002). Learning covert phonological interaction: an analysis of the problem posed by the interaction of stress and epenthesis. Report no. RuCCS-TR-72. Piscataway, NJ: Rutgers Center for Cognitive Science. (ROA 543)

Baković, Eric (1999). Deletion, insertion, and symmetrical identity. *Harvard Working Papers in Linguistics* **7** (ROA 300).

Baković, Eric (2000). *Harmony, dominance, and control*. PhD dissertation, Rutgers University, New Brunswick, NJ (ROA 360).

Benua, Laura (1997). *Transderivational identity: phonological relations between words*. PhD dissertation, University of Massachusetts, Amherst.

Bermúdez-Otero, R. (1999a) Opacity: evidence from West Germanic Gemination. Handout of paper presented at the Spring Meeting of the Linguistics Association of Great Britain, Manchester, 10 April 1999. Available at: www.bermudez-otero.com/lagb.pdf

Bermúdez-Otero, R. (1999b). *Constraint interaction in language change* [*Opacity and globality in phonological change.*] PhD dissertation, University of Manchester / Universidad de Santiago de Compostela. Available at: www.bermudez-otero.com/PhD.pdf

Bermúdez-Otero, R. (2001). Voicing and continuancy in Catalan: a nonvacuous Duke-of-York gambit and a Richness-of-the-Base paradox. Unpublished manuscript: University of Manchester, November 2001. Available at: www.bermudez-otero.com/Catalan.pdf

Bermúdez-Otero, R. (2002a) Analogy in stratal OT. Handout of paper presented at the Conference on English phonology, Toulouse, 26 June 2002. Available at: www.bermudez-otero.com/toulouse2002.pdf

Bermúdez-Otero, R. (2002b) Cyclicity or sympathy? A case study. Handout of paper presented at the 10th Manchester Phonology Meeting, Manchester, 25 May 2002. Available at: www.bermudez-otero.com/10mfm.pdf

Bermúdez-Otero, R. (in preparation). *The life cycle of constraint rankings: studies in early English morphophonology*. Outline available at: www.bermudez-otero.com/lifecycle.htm

Bermúdez-Otero, R. & R. M. Hogg (forthcoming). The actuation problem in Optimality Theory: phonologization, rule inversion, and rule loss. In D. E. Holt (ed), *Optimality Theory and language change.* Dordrecht: Kluwer.

Bermúdez-Otero, R. & A. McMahon (in preparation). English phonology and morphology. In B. Aarts & A. McMahon (eds), *The handbook of English linguistics*. Oxford: Blackwell.

Bromberger, S. & M. Halle (1989). Why phonology is different. *Linguistic Inquiry* **20**: 51-70.

Chambers, J. K. (1973). Canadian raising. *Canadian Journal of Linguistics* **18**: 113-35.

Chomsky, N. (1964). Current issues in linguistic theory. In J. Fodor & J. Katz (eds), *The structure of language.* Englewood Cliffs, NJ: Prentice Hall. 50-118.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1986). *Knowledge of language: its nature, origin, and use*. Westport, Conn.: Praeger.

Chomsky, N. (2001). Beyond explanatory adequacy. *MIT Occasional Papers in Linguistics* **20**.

Chomsky, N. & M. Halle (1968). *The sound pattern of English*. New York: Harper and Row.

de Lacy, P. (1998). Sympathetic stress. Ms., University of Massachusetts, Amherst. (ROA 294)

Dinnsen, D. A., L. W. McGarrity, K. M. O'Connor & K. A. B. Swanson (2000). On the role of sympathy in acquisition. *Language Acquisition* **8**: 321-361.

Gussenhoven, C. (1986). English plosive allophones and ambisyllabicity. *Gramma* **10**: 119-141.

Hale, K. (1973). Deep-surface canonical disparities in relation to analysis and change: an Australian case. In T. A. Sebeok (ed), *Current trends in linguistics*. Vol. 11: *Diachronic, areal, and typological linguistic*s. The Hague: Mouton. 401-458.

Halle, M. & W. J. Idsardi (2000). Stress and length in Hixkaryana. *The Linguistic Review* **17**: 199-218.

Hayes, B. (1999). Phonological acquisition in Optimality Theory: the early stages. Ms, UCLA (ROA 327). To appear in R. Kager, J. Pater & W. Zonneveld (eds), *Fixing priorities: constraints in phonological acquisition*. Cambridge: Cambridge University Press.

Itô, J. & R. A. Mester (1997). Sympathy Theory and German truncations. In V. Miglio & B. Morén (eds), *Proceedings of the Hopkins Optimality Workshop/Maryland Mayfest 1997*. University of Maryland Working Papers in Linguistics **5**: 117-139.

Inkelas, S. (1995). The consequences of optimization for underspecification. *Proceedings of the North East Linguistic Society* **25**: 287-302.

Inkelas, S. (2000). Phonotactic blocking through structural immunity. In B. Stiebels & D. Wunderlich (eds), *Lexicon in focus*. (Studia Grammatica 45). Berlin: Akademie Verlag. 7-40.

Jensen, J. T. (2000). Against ambisyllabicity. *Phonology* **17**: 187-235.

Joos, M. (1942). A phonological dilemma in Canadian English. *Language* **18**: 141-4.

Kahn, D. (1976). *Syllable-based generalizations in English phonology*. PhD dissertation, MIT.

Kaisse, E. M. & P. A. Shaw (1985). On the theory of Lexical Phonology. *Phonology Yearbook* **2**: 1-30.

Kenstowicz, M. (1994). *Phonology in generative grammar*. Cambridge, MA: Blackwell.

Kiparsky, P. (1979). Metrical structure assignment is cyclic. *Linguistic Inquiry* **10**: 421-441.

Kiparsky, P. (1993). Blocking in nonderived environments. In Hargus, S. & E. M. Kaisse (eds), *Studies in Lexical Phonology*. (Phonetics and Phonology **4**). San Diego: Academic Press. 277-313.

Kiparsky, P. (2000). Opacity and cyclicity. *The Linguistic Review* **17**: 351-365.

Kiparsky, P. (forthcoming). *Paradigm effects and opacity*. Stanford: CSLI Publications.

Kisseberth, C. (1970). On the functional unity of phonological rules. *Linguistic Inquiry* **1**: 291-306.

Kuryłowicz, J. (1949). La nature des procès dites 'analogiques'. *Acta Linguistica* **5**: 15-37.

Lahiri, Aditi (1982). *Theoretical implications of analogical change: evidence from Germanic languages*. PhD dissertation, Brown University.

Lahiri, Aditi & B. Elan Dresher (1983-84). Diachronic and synchronic implications of declension shifts. *The Linguistic Review* 3: 141-163.

Mańczak, W. (1958). Tendences générales des changements analogiques. *Lingua* **7**: 298-325, 387-420.

McCarthy, J. J. (1999). Sympathy and phonological opacity. *Phonology* **16**: 331-399.

McCarthy, J. J. (2003). Sympathy, cumulativity, and the Duke-of-York gambit. In C. Féry & R. v. d. Vijver (eds), *The syllable in Optimality Theory*. Cambridge: Cambridge University Press.

Minkova, D. & R. Stockwell (forthcoming). English vowel shifts and 'optimal' diphthongs: is there a logical link. In D. E. Holt (ed), *Optimality Theory and language change*. Dordrecht: Kluwer.

Pater, J. (1995). Austronesian nasal substitution and other NC effects. Ms., McGill University (ROA 92). Revised version published in R. Kager, H. van der Hulst & W. Zonneveld (eds), *The prosody-morphology interface*. Cambridge: Cambridge University Press. 310-343.

Prince, A. & P. Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Report no. RuCCS-TR-2. New Brunswick, NJ: Rutgers University Center for Cognitive Science. (Revised August 2002: ROA 537.)

Prince, A. & B. Tesar (1999). Learning phonotactic distributions. Report no. RuCCS-TR-54. New Brunswick, NJ: Rutgers University Center for Cognitive Science. (ROA 353)

Tesar, B. & P. Smolensky (2000). *Learnability in Optimality Theory.* Cambridge, MA: MIT Press.

Vance, T. (1982). 'Canadian Raising' in some dialects of the Northern United States. *American Speech* **62**: 195-210.

Wells, J. C. (1982). *Accents of English*. Vol. 3: *Beyond the British Isles*. Cambridge: Cambridge University Press.

Yip, M. (1996). Lexicon optimization in languages without alternations. In J. Durand & B. (eds). *Current trends in phonology: models and methods*. Vol. 2. Salford: European Studies Research Institute. 759-790.