In this way, candidate chain theory restricts OT's power to perform certain global optimizations. This is also a major point of difference between chain theory and standard derivational models, which have no analogous notion of improvement or progress in their derivations (see §3.2.4.3).

The final chapter of this book presents case studies of two languages with significant amounts of opaque phonology. Because of my background and knowledge, both of the languages are varieties of Arabic, but I believe that they are fully representative of the kinds of opaque interactions that can be found in other, unrelated languages. Furthermore, most of the processes that are discussed are independent innovations rather than the legacies of a common ancestor. These case studies are a necessary adjunct to the theoretical proposal because many of the more complex issues in studying opacity only arise in analyses of sufficient depth to show more than two processes interacting (cf. Cathey and Demers 1970).

1.3 How should this book be read?

Readers who are new to OT are advised not to start with this book and to begin instead with a textbook introduction like Kager (1999a) or an overview like McCarthy (2002b). Either of those works will provide more than enough background to understand and critically evaluate the contents of the following chapters. The majority of readers will probably want to proceed linearly through §2 and §3, and then sample the extended analyses in §4. Those who are familiar with previous work on opacity before and since OT could skim rather than read §2. Readers who prefer praxis to theory may want to try reading §4 on the basis of just the brief introduction to candidate chains in §1.2, but I would not recommend it.

Notes

1 The *mankind* example comes from Donegan and Stampe (1979: 148–149). They also describe many other opaque interactions in English casual speech and dialect variation.

2 Throughout, I use the expression 'classic OT' as shorthand for the approach that has become a *de facto* standard, a synthesis of Prince and Smolensky's (2004) original proposals with correspondence-based faithfulness (McCarthy and Prince 1995, 1999). Among the characteristics of classic OT is a universal, finite constraint component Con that is limited to markedness and faithfulness constraints, and a single Eval that evaluates fully formed output candidates that show the effects of all phonological processes in parallel.

2 Opacity, derivations, and Optimality Theory

2.1 Overview

This chapter begins (§2.2) by explaining what opacity is and how it is analyzed in rule-based phonology. The discussion then turns (§2.3) to a description of 'classic' Optimality Theory, the problems that opacity presents for classic OT, and various ideas about how to modify the classic theory to accommodate it. The conclusion I draw (§2.4) is that there is something fundamentally correct about rule-based phonology's serial derivation, leading to the proposal in §3 for an analogue of the serial derivation in a framework that retains all of classic OT's essential elements.

2.2 Opacity and derivations

2.2.1 Levels of representation

The theory of generative phonology recognizes two principal levels of representation, underlying and surface. At the underlying level, every morpheme has a unique representation. For example, the three principal surface alternants of the English plural suffix — [-z], [-s], and [-əz] — are derived by phonological rules from a single underlying representation, such as /-z/. Only suppletive or allomorphic alternants of morphemes require distinct underlying representations, such as the plural allomorphs /-ən/ of *children* and /-iː-/ of *geese*.

When a morpheme alternates nonsuppletively, its underlying representation must be discovered by the analyst and by the learner. In paradigms like German [bʊnt]/[bʊntə] 'multicolored/pl.' and [bʊnt]/[bʊndə] 'federation/pl.', distinct underlying representations are required because there are distinct patterns of voicing alternation: /bʊnt/ 'multi-colored' is voiceless throughout its paradigm and /bʊnd/ 'federation' alternates between voiced and voiceless. In theory and in actual practice, the relationship between the hypothesized

underlying representation and the observed surface paradigm is sometimes less transparent than this.

Some recent research explores alternatives to positing an underlying level of representation. These approaches are monostratal in the sense that they recognize only a single level of representation, the surface form. In Declarative Phonology (Scobbie, Coleman, and Bird 1996), the work of underlying representations is done by constraints that describe morphemes. These descriptions are crucially incomplete in the case of alternating morphemes: e.g., for German [bʊnt]/[bʊndə], a constraint requires a final alveolar stop in 'federation' but says nothing about its voicing. Another monostratal approach seeks to express phonological generalizations purely in terms of relations between surface forms (e.g., Albright 2002, Burzio 2002).

In this context, it is worth reviewing the reasons why generative phonology posits an underlying level of representation (see Kenstowicz and Kisseberth 1979: chapter 6 for an overview of the evidence). The main argument comes from paradigms where the relationships among surface forms make sense only when mediated by an underlying form that is distinct from all of the surface forms. Schane's (1974) Palauan example in (2-1) is a well-known case. Because unstressed vowels reduce to [ə] and there is only one stress per word, disyllabic roots like 'cover an opening' and 'pull out' never show up with more than one surface nonschwa vowel. The hypothesized underlying representations /daŋob/ and /teʔib/ record the quality of the vowels as they appear when stressed in different members of the paradigm. These underlying representations incorporate all of the unpredictable phonological information about these morphemes. In generative phonology, the underlying representation of a root is the nexus of a set of related words, so it must contain sufficient information to allow the surface forms of all of those words to be derived by the grammar of the language. (See §4.3.3 for detailed argumentation in support of underlying representations in a case similar to Palauan.)

(2-1) Palauan Vowel Reduction

Underlying	Present Middle	Future Participle	
/daŋob/	mə-ˈdaŋəb	dəˈŋob-l	'cover an opening'
/teʔib/	mə-ˈteʔəb	təˈʔib-l	'pull out'

Generative phonology in the tradition of *The Sound Pattern of English* (SPE — Chomsky and Halle 1968) also allows for any number of levels intermediate between the underlying and surface levels. These intermediate levels are the result of sequential application of phonological rules. If a language has n rules in its grammar, it has $n-1$ intermediate representations, each of which is a potentially distinct way of representing the linguistic form that is being derived.

In Palauan, for example, the *SPE* theory requires an intermediate level at which stress has been assigned but vowel reduction has not yet applied: /daŋob-l/ $\rightarrow_{\text{stress}}$ [daˈŋobl] $\rightarrow_{\text{reduction}}$ [dəˈŋobl]. Indeed, *SPE* requires rules to apply sequentially even when simultaneous application would produce the same result.

2.2.2 Derivations

Any mapping from the underlying to the surface level of representation is a derivation. In this sense, any multistratal theory of phonology is derivational, including classic OT. The various multistratal theories differ significantly, however, in the complexity and internal organization of the derivations they posit.

The *SPE* approach to derivations retains considerable currency because it is often assumed even in contemporary research that has moved far beyond *SPE*'s other hypotheses about rules and representations (see §2.2.6). In *SPE*, the grammar consists of an ordered list of rules. The rules are applied in a strict sequence, with the output of rule i supplying the input to rule i+1. The output of each rule (except the last) is therefore a level of representation intermediate between the underlying and surface levels.

An important insight, due originally to Kiparsky (1968), is that rules may have different functional relationships to one another. In the least interesting case, a pair of rules may not interact at all — an example would be word-initial vowel epenthesis and word-final obstruent devoicing. When rules do interact, however, the functional relationship between them can often be classified as feeding or bleeding.

Rule A is said to *feed* rule B if A can create additional inputs to B. If A in fact precedes B, then A and B are in feeding order. (If B precedes A, then they are in counterfeeding order, which will be explained in §2.2.3.) An example of feeding order is the interaction between vowel and consonant epenthesis in Classical Arabic. Words that begin with consonant clusters receive prothetic [ʔi] (or [ʔu], if the next vowel is also [u]). As the derivation in (2-2) shows, prothesis of [ʔi] is the result of a feeding interaction between [i] epenthesis before word-initial clusters (= rule A) and [ʔ] epenthesis before word-initial vowels (= rule B).

(2-2) Feeding order in Classical Arabic

Underlying	/dˤrib/	'beat (m. sg.)!'
Vowel epenthesis	idˤrib	
[ʔ] epenthesis	ʔidˤrib	
Surface	[ʔidˤrib]	

Rule A is said to *bleed* rule B if A can eliminate potential inputs to B. If A in fact precedes B, then A and B are in bleeding order. (If B precedes A, then they are in counterbleeding order, which will also be explained in §2.2.3.) For example, in a southern Palestinian variety of Arabic, progressive assimilation of pharyngealization (= rule B) is blocked by high front segments, among them [i]. When the vowel [i] is epenthesized into triconsonantal clusters (= rule A), it also blocks assimilation, as shown in (2-3) (Davis 1995).

(2-3) Bleeding order in southern Palestinian Arabic

Underlying	/batˤn-ha/ 'her stomach'
Vowel epenthesis	batˤinha
Progressive assimilation	*Blocked*
Regressive assimilation	bˤaˤtˤinha
Surface	[bˤaˤtˤinha]

Feeding and bleeding orders have something in common: when rules apply in feeding or bleeding order, those structures that are derived by rules are treated exactly the same as similar structures that were already present in underlying representation. For example, the process of [ʔ]-epenthesis in Classical Arabic applies to words with an underlying initial vowel, /al-walad-u/ → [ʔalwaladu] 'the boy (nominative)', and also to words with a derived initial vowel, such as the intermediate representation [idˤrib] in (2-2). Likewise, epenthetic and nonepenthetic [i] equally block progressive assimilation in Palestinian Arabic, as shown by (2-3) and /sˤiħħa/ → [sˤiħħa], *[sˤiˤħˤħˤaˤ] 'health'. In feeding and bleeding interactions, what you see is what you get: when derived and underived structures are identical, they exhibit identical phonological behavior. This is emphatically not the case with counterfeeding and counterbleeding interactions.

2.2.3 Opacity in derivations

If rule A feeds rule B and they are applied in the order B precedes A, then these rules are said to be in *counterfeeding* order. For example, in a Bedouin Arabic dialect (see §4.3.3), there are processes raising short /a/ to a high vowel in a nonfinal open syllable (= rule A) and deleting short high vowels in nonfinal open syllables (= rule B). These processes are in a feeding relationship, since raising has the potential to create new inputs to deletion. But their order is actually counterfeeding, as shown in (2-4). High vowels derived by raising are treated differently from underlying high vowels; only the underlying high vowels are subject to deletion. When rules apply in feeding order, derived and underlying structures behave alike, but when they apply in counterfeeding order, derived and underlying structures behave differently.

(2-4) Counterfeeding order in Bedouin Arabic[1]

Underlying	a. /dafaʔ/ 'he pushed'	b. /ʃarib-at/ 'she drank'
Deletion	—	ʃarbat
Raising	difaʔ	—
Surface	[difaʔ]	[ʃarbat]

The same is true of *counterbleeding* order, where rule A bleeds rule B but they are applied with B preceding A. In this same Arabic dialect, there is also a process palatalizing velars when they precede front vowels (see §3.3.3). Deletion (= rule A) bleeds palatalization (= rule B), since deletion can remove a high front vowel that would condition velar palatalization. But their order is counterbleeding, as shown in (2-5). High front vowels, even when they are absent from surface forms, induce adjoining velars to palatalize. Effects like this are typical with counterbleeding order.

(2-5) Counterbleeding order in Bedouin Arabic

Underlying	a. /ħaːkim-iːn/	b. /t-ħakum-in/
Palatalization	ħaːkʲimiːn	—
Deletion	ħaːkʲmiːn	tħakmin
Surface	[ħaːkʲmiːn]	[tħakmin]
	'ruling (masculine plural)'	'they (feminine) rule'

The result of counterfeeding and counterbleeding interactions is phonological opacity. Kiparsky's (1973: 79, 1976: 178–179) definition of opacity appears in (2-6). Clause (c) of this definition describes all processes of neutralization and so it is not relevant to our concerns here. We will therefore focus on clauses (a) and (b).

(2-6) Opacity

A phonological rule P of the form $A \rightarrow B / C\_\_D$ is *opaque* if there are surface structures with any of the following characteristics:

a. instances of A in the environment $C\_\_D$,
b. instances of B derived by P that occur in environments other than $C\_\_D$,
or c. instances of B not derived by P that occur in the environment $C\_\_D$.

In the derivation /dafaʔ/ → [difaʔ] in (2-4), the high-vowel deletion rule is opaque under clause (a) of this definition: [difaʔ] has [i] (= *A*) in an open syllable (= *C__D*). Rules applied in counterfeeding order produce opacity of the clause (a) type, in which surface forms contain phonological structures that look like they should have undergone some process but in fact did not.[2]

In the derivation /ħaːkim-iːn/ → [ħaːkʲmiːn] in (2-5), the palatalization rule is opaque under clause (b) of this definition: [ħaːkʲmiːn] has [kʲ] (= *B*) derived

by palatalization (= *P*), but [kʲ] is not adjacent to a front vowel (= *C__D*). Rules applied in counterbleeding order produce opacity of this type, in which surface forms contain derived phonological structures without the context that shows how they were derived.

Counterfeeding and counterbleeding interactions supply the best — arguably, the only — evidence for language-particular rule ordering. It is not surprising, then, that skepticism about stipulated, language-particular ordering stimulated efforts to deny that opaque interactions involve living phonological processes (cf. §1.1). According to the proponents of Natural Generative Phonology (NGP), authentic phonological rules must state surface-true generalizations and they must be unordered (Hooper [Bybee] 1976, 1979, Vennemann 1972, 1974). NGP therefore maintains that opaque processes are merely the lexicalized residue of sound changes that are no longer productive — opaque rules were said to be 'not psychologically real'. (Recent work advocating similar views in an OT context includes Green (2004), Mielke, Hume, and Armstrong (2003), and Sanders (2002, 2003).) In fact, much if not all of the abstractness controversy of the 1970's, which dealt with proposed limits on the degree of disparity between underlying and surface representations (see Kenstowicz and Kisseberth 1977: Chapter 1, 1979: Chapter 6), was really an argument about opacity, since abstract underlying forms can influence the output only if opaque rules apply to them.

Certainly, there have been dubious analyses based on opaque rules and excessively abstract underlying forms, but outright denial of all opaque interactions is an empirically unsupportable overreaction. The example of Bedouin Arabic is instructive. (See §4.3.3 for detailed discussion.) Al-Mozainy (1981) presents several arguments that the opaque processes in this language are alive and productive. First, they are active in borrowed words. Second, high vowel deletion, even though it is opaque, applies productively in external sandhi, as shown in (2-7). If a process applies in external sandhi, it cannot be lexicalized, since it is impossible to list the infinite number of word collocations that the syntax provides.[3]

(2-7) Phrase-level deletion in Bedouin Arabic (Al-Mozainy 1981: 50–51)

/kaːtib al-ʒawaːb/	kaːt.bal.ʒu.waːb	'writing the letter'
	*kaː.ti.bal.ʒu.waːb	
/tiʔtˤuːnih al-museːʔiːdi/	tiʔ.tˤuːn.hal.m.seː.ʔiː.di	'you give it to the one from the clan of Musaiʿīd'
	*tiʔ.tˤuː.ni.hal.m.seː.ʔiː.di	

Third, the most compelling evidence that raising is productive comes from a kind of play language. Although raising usually affects any short /a/ in a nonfinal open syllable, there are phonological conditions under which raising regularly fails to apply: after a guttural consonant ([ʔ], [h], [ʕ], [ħ], [χ], [ʁ]), or before a guttural consonant or coronal sonorant ([l], [r], [n]) that is itself followed by [a]. Bedouin Arabic has a secret language that permutes the consonants of the root, and this will sometimes affect the position of gutturals or coronal sonorants relative to the potentially raised vowel. When that happens, the vowel raises or fails to raise in exact conformity with these generalizations, as (2-8) shows. Other secret language data show that palatalization is also productive, even though it is opaque (see §3.3.3). In sum, the opaque phonology of Bedouin Arabic is also its living, productive phonology. (For further examples of processes that are productive yet opaque, see Donegan and Stampe (1979).)

(2-8) Raising alternations in a secret language

/dafaʔ/	Underlying representation
difaʔ	Unpermuted form
fidaʔ	Raising as expected
daʔaf	No raising before guttural + [a]
faʔad	"
ʔadaf	No raising after guttural
ʔafad	"

Although this sort of evidence shows that opacity is a fact of phonological life, certain types of opacity have received and deserve a skeptical reception. A famous example is *SPE*'s /ɹixt/ → [ɹɑjt] *right*. The point is that a few dubious analyses are not grounds to reject a theoretical construct, particularly when it is strongly supported by sound analyses, as it is in Bedouin Arabic.

A type of opacity that received particular attention in the 1970's is the Duke-of-York derivation (Hogg 1978, Pullum 1976). Like the eponymous Duke of the nursery rhyme,[4] underlying /A/ is changed by a rule to intermediate [B], but a later rule changes [B] back into [A]. Unlike the Duke's peregrinations, this activity is not as pointless as it seems: during the temporary [B] stage, erstwhile /A/ may opaquely escape an A-affecting process or cause a B-triggered one. More often, though, Duke-of-York derivations are simply an artifact of the commitment to sequential rule application. We will return to this topic, with exemplification, in §2.3.2.

2.2.4 Simultaneous application

Discussions of rule ordering often overlook an important alternative to the sequential derivation: simultaneous application. In many cases, rules could be applied simultaneously with no loss of generality, and so it is worth exploring which phenomena are and are not consistent with simultaneous application (for

discussion, see Anderson 1974: 64–67, Donegan and Stampe 1979: 150, Hyman 1993: 204ff., Koutsoudas 1976, Koutsoudas, Sanders, and Noll 1974: 5–8).

Since simultaneous application is a somewhat unfamiliar notion, we should first get clear on what it means in rule-based phonology. A phonological rule describes a configuration that must be met in the rule's input — the rule's structural description — and a change that is to be effected in the rule's output — its structural change. If two rules are applied simultaneously, then their structural descriptions are analyzing exactly the same representation. It follows, then, that neither rule has access to any information that is contributed by the other rule's structural change. In sequential application, by contrast, the later rule always has access to information contributed by the earlier rule's structural change.

Opaque interactions are often compatible with simultaneous application, but transparent interactions require sequential application. The counterbleeding derivation /ħa:kim-i:n/ → [ħa:kʲmi:n] in (2-5), for example, would also work if the rules of palatalization and deletion were applied simultaneously. The structural description of the palatalization rule analyzes an input that contains [k] before [i], and so [k] is palatalized with complete indifference to the fact that the deletion rule is analyzing that same input toward the goal of deleting [i]. The important thing in this opaque derivation is that deletion must not precede palatalization; that desideratum could in principle be fulfilled by ordering palatalization before deletion, as in (2-5), or by requiring them to apply simultaneously.

Similarly, the counterfeeding derivation /dafaʔ/ → [difaʔ] in (2-4) is possible if deletion of high vowels and raising of low vowels apply simultaneously. The structural description of the high-vowel deletion rule is not met by /dafaʔ/, but the raising rule's structural description is met, so only raising actually applies. The important thing in this opaque derivation is that deletion should not apply to the output of raising; that desideratum could in principle be fulfilled by ordering raising before deletion, as in (2-4), or by requiring them to apply simultaneously.

Feeding and bleeding interactions, however, are incompatible with simultaneous application. In the feeding derivation /dˤrib/ → [ʔidˤrib] (2-2), for instance, the structural description of [ʔ] epenthesis is not met until after vowel epenthesis has applied, so sequential application is necessary. In the bleeding derivation /batˤn-ha/ → [bˤaˤtˤinha] (2-3), simultaneous application of vowel epenthesis and progressive assimilation would produce the result *[bˤaˤtˤinˤhˤaˤ], in which the epenthetic vowel is neither subject to nor a blocker of assimilation.

It is interesting that simultaneous application of rules typically produces opaque interactions but not transparent ones (unless the rules do not interact at

it evaluates candidates in which the effects of several
taneously, can model transparent interactions but not
:). The reason for this difference is that rules and OT
nalyze different levels of representation. The structural
et by the rule's *input*, which is sometimes identical to
ation. The structural description of an OT markedness
ultimate *output*, the surface representation. Opacity
onditions obtaining in presurface representations,
equires reference to conditions obtaining in surface

lering

the rules are applied is *extrinsic*, which means that it
the language-particular grammar and cannot usually
orm or function. From about 1969 through 1980, a
eloped around the question of whether some or even
g could be predicted. (See Anderson (1979: 15–18)
ief surveys or Anderson (1974) and Kenstowicz and
rs 4, 6) for more extensive discussion.)
ogy of Classical Arabic must include a statement to
nthesis precedes [ʔ] epenthesis to ensure that these
served in (2-2). In some revisions of that model (e.g.,
las, Sanders, and Noll 1974), this ordering statement
ous on the grounds that feeding order is unmarked
s feeding order natural? If rules are allowed to apply
derivation when their structural descriptions are
e the same as (2-2). Feeding orders maximize rule
d in §2.2.2, feeding orders also help to ensure that
zations about surface structure: in Arabic, no word
e [ʔ] epenthesis is ordered after and thereby fed by
erminology of Donegan and Stampe (1979: 157),
owel epenthesis and [ʔ] epenthesis would act as a
reventing it from acting on derived representations
enthesis generalization not categorically true.
r was generally seen as natural and a second natural
t, there is disagreement in the literature of that era
er the other natural rule order is bleeding or coun-
work on this topic, Kiparsky (1968) argued that
aximize feeding and minimize bleeding orders. On

the assumption the languages are attracted toward natural rule orders, this would mean that feeding and counterbleeding orders are natural. Anderson (1974) integrates this idea into his theory of local ordering, according to which feeding and counterbleeding order constitute a default case that can only be overridden by language-particular stipulation. Anderson's evidence includes analyses, none of them uncontroversial, in which maintaining the natural interaction between a pair of rules can cause them to apply in different orders within a single language. (This is the sense in which ordering is 'local': the theory comprehends ordering as a local relation between a pair of rules rather than a global list of ordered rules in the *SPE* fashion. Cf. §3.2.3.) Koutsoudas, Sanders, and Noll (1974) also argue for the naturalness of counterbleeding order.

Another body of work took the position that bleeding rather than counterbleeding order is natural (Iverson 1974, Kenstowicz and Kisseberth 1971, Kiparsky 1971). Apart from disagreements about analyses, the dispute is really one about the principle that determines the natural orders. If feeding and counterbleeding orders are natural, then what makes them natural is a principle that favors maximizing rule applicability: if A feeds B, then A supplies additional opportunities for rules to apply; and if A is not allowed to bleed B, then A cannot steal away some of B's opportunities to apply. If feeding and bleeding orders are natural, then what makes them natural is a principle that favors maximizing rule transparency: counterfeeding and counterbleeding orders produce opacity, whereas feeding and bleeding orders produce transparency, in which the effects of phonological generalizations are visible at surface structure.

In the course of research during the 1970's, these and other ordering principles were discussed, and there were even proposals about priority relationships among them (Anderson 1974: 217–218, Iverson 1976). The ultimate goal of the research program, according to some (e.g., Koutsoudas, Sanders, and Noll 1974), was the elimination of all language-particular ordering statements in favor of universal principles of applicational precedence. Supposedly *prima facie* arguments against this position have been adduced, such as two Canadian English dialects that differ solely in rule order (Bromberger and Halle 1989, Joos 1942), but in reality the argument is not that easy to make (Iverson 1995: 612–613). There never was a knock-down argument in support of language-particular ordering, nor was there general agreement on rule-ordering principles. Instead, the decade ended with a tacit consensus that research on universals of rule ordering had gone about as far as it could go.

2.2.6 Later developments

Interest in the topic of rule ordering waned around 1980. (An important exception is Goldsmith (1993b).) As the focus of phonological research moved elsewhere, however, matters of rule ordering and interaction sometimes reemerged in new contexts.

The development of nonlinear phonology and underspecification theory, beginning with works like Goldsmith (1976a), Kahn (1976), Liberman (1975), Liberman and Prince (1977), Clements and Ford (1979), McCarthy (1981), Prince (1983), and Archangeli (1984), took some of the analytic pressure off of phonological rules and shifted it to well-formedness constraints on phonological representations. In principle, an enriched theory of representations might lead to a reduction in the need for language-particular rule ordering, but in actual practice this line of research received little attention.

Satisfaction of representational constraints, however, required a new class of persistent rules that apply automatically at any point in the derivation when they are required (Chafe 1968, Myers 1991a). For example, when a consonant becomes unsyllabified in the course of a derivation, a persistent rule immediately adjoins it to a nearby syllable: /patika/ → $[pa]_\sigma$ $[ti]_\sigma$ $[ka]_\sigma$ → $_{\text{syncope}}$ $[pa]_\sigma$ *t* $[ka]_\sigma$ →$_{\text{persistent}}$ $[pat]_\sigma$ $[ka]_\sigma$. An important role of persistent rules, then, is to repair violations of well-formedness conditions, thereby ensuring that these conditions are respected not only at the beginning or end of the derivation but also in the middle. The free (re-)applicability of persistent rules is, of course, consistent with the principle favoring maximal rule application that was mentioned at the end of §2.2.5.

Another relevant post-1980 development is the theory of Lexical Phonology (Kaisse and Hargus 1993b, Kaisse and Shaw 1985, Kiparsky 1982, 1985, Mohanan 1982, among many others). Lexical Phonology is an extension of *SPE*'s theory of cyclic rule application. Certain rules may be designated as cyclic — in *SPE*, these are the English stress rules — and this causes them to apply repeatedly to successively larger morphological or syntactic constituents. The cycle accounts for transderivational similarities like the following:[5]

(i) Monomorphemic words like ˌ*Kalama*ˈ*zoo* and ˌ*Winnepe*ˈ*saukee* exhibit the normal English stress pattern when three light syllables precede the main stress. Derived words like *ac*ˌ*credi*ˈ*tation* and *i*ˌ*magi*ˈ*nation* deviate from this pattern under the influence of *ac*ˈ*credit* and *i*ˈ*magine*.

(ii) A closed, sonorant-final syllable is normally unstressed in prestress position: ˌ*seren*ˈ*dipity*, ˌ*gorgon*ˈ*zola*, ˌ*Pennsyl*ˈ*vania*. But the same kind of syllable may be stressed in the derived words ˌ*au*ˌ*then*ˈ*ticity* and ˌ*con*ˌ*dem*ˈ*nation* under the influence of ˌ*au*ˈ*thentic* and *con*ˈ*demn*.

In *SPE*, the aberrant stress of derived words is explained by their bracketing and cyclic application of stress. The stress rules first apply on the inner constituents of [*accredit*]*ation* or [*authentic*]*ity* and then on the outer constituents. The primary stress assigned on the first cycle becomes a secondary stress on the second cycle, when the stress rule reapplies and a new primary stress is assigned further to the right. Monomorphemic *Kalamazoo* and *serendipity* have no inner cycle, so they show the effects of just a single pass through the stress rules.

Lexical Phonology departs from *SPE* in regarding cyclic application as the norm rather than the exception for certain phonological rules. In addition, Lexical Phonology imposes further structure on the grammar, dividing the phonology up into separate components, called strata. At a minimum, there are two such strata, lexical and postlexical. The input to the lexical stratum is the underlying representation; the output of the lexical stratum is the input to the postlexical stratum; and the output of the postlexical stratum is the surface representation. Each stratum is a separate phonological grammar, though specific overlap requirements have sometimes been imposed (Borowsky 1986, Kiparsky 1984: 141–143, Myers 1991b) (see §2.3.4.2). It is usually assumed that the lexical stratum actually consists of several strata, and at each lexical stratum a different set of morphological and phonological processes may be in effect. For example, English suffixes like *-ity* are affixed in the first lexical stratum, and that is also where the stress assignment rules apply. Suffixes like *-ness* are not attached until the second lexical stratum, at which point the stress assignment rules are no longer active. That is why suffixes like *-ity* are stress-determining and suffixes like *-ness* are stress-neutral. It is sometimes also assumed that rules apply cyclically within each lexical stratum, as each affix of that stratum is added: e.g., /period/ → *'period* → *peri'odic* → *perio'dicity*, all within the first lexical stratum.

Lexical Phonology retains *SPE*'s assumption that the rules within a grammar (= a stratum) are in a strict linear order. Despite this within-grammar strict ordering, the same rule can be observed to reapply at different points in the course of an entire derivation. As in *SPE*, the cycle offers one opportunity: a rule can reapply in the same stratum as multiple affixes are added. But even without any affixation at all, a rule can reapply if it is assigned to more than one stratum. (This situation is not unusual.) If a rule is included in the grammar of more than one stratum, it will simply reapply when the later stratum is reached. Lexical Phonology is thereby able to reanalyze some (perhaps all) of the evidence that had earlier been adduced in favor of a principle of unmarked feeding order or maximizing rule application. An example is Kiparsky's (1984) reanalysis of the interaction of Icelandic *u*-umlaut and syncope, which had previously been

cited by Anderson (1974) as evidence for local ordering. Instead of allowing *u*-umlaut and syncope to apply in either order, whichever produces a feeding relationship, the Lexical Phonology approach fixes the within-stratum order as *u*-umlaut precedes syncope, but then allows *u*-umlaut to follow syncope by reapplying in a later stratum.

Assignment of rules to different strata offers a way of imposing extrinsic ordering on them: if rule A applies only in stratum 1 and rule B applies only in stratum 2, then A necessarily precedes B. Therefore, assignment of rules to strata could be used to reproduce some of the effects of *SPE*-style extrinsic ordering. This leads to some questions: Is extrinsic ordering within strata truly necessary? Could all rules in the same stratum apply simultaneously or in a universally predictable order? These questions were not asked, much less answered, in mainstream work on Lexical Phonology, though they were discussed in work that is not usually identified with the Lexical Phonology research program (Goldsmith 1993a, Lakoff 1993). In any case, the questions persist to this day, as we will see in §2.3.4.2.

Apart from these developments, the common consensus about rule ordering and opacity did not change very much in the period after 1980. Most phonologists, perhaps more from a lack of interest than strong conviction, continued to assume something like the *SPE* model of rule interaction.

2.3 Opacity in Optimality Theory

2.3.1 Properties of classic OT

This section is not intended as an introduction to or comprehensive overview of OT (for the former see Kager (1999a), and for the latter see Prince and Smolensky (2004) or McCarthy (2002b)). Rather, the goal is to review those aspects of OT that assume particular significance in the analysis of opacity.

In OT, a grammar of a language is a ranking of constraints. Ranking differs from language to language, so the ranking relation between any given pair of constraints is not generally predictable. Because language-particular ranking offers a way of accounting for language differences, it is reasonable (though not strictly necessary) to adopt the null hypothesis that the constraints themselves are universal and so are drawn from a universal constraint component, called Con. If Con is indeed universal, as is standardly assumed in classic OT, then it is fair to say that OT is an inherently typological theory of language. In other words, OT and a specific hypothesis about Con combine to predict all and only the possible grammars of human languages.