Derivations and levels of representation

Comp. by: parvathy Date:13/7/06 Time:15:55:04 Stage:First Proof File Path://Spsind002s/
Production/PRODENV/0000000009/0000000186/0000000005/0000059676.3D Proof by:
QC by: Author: McCarthy

5

Derivations and levels of representation

John J. McCarthy

5.1 Introduction

In the theory of generative phonology, the phonological grammar of a language is regarded as a function from underlying to surface forms: /kæt+z/ → [kæts] 'cats'. Underlying and surface form are known as *levels of representation*, and the mapping between them is a *derivation*. This chapter describes the rationale for positing distinct levels of representation, various views of how many and what kind of levels of representation there are, and the nature of the derivations that link different levels of representation.

5.2 Levels of representation

In structuralist phonology of the first half of the twentieth century (see Joos 1957 for many examples), three levels of representation were recognized. One level, called *allophonic* or *phonetic*, offers a more or less accurate transcription of the actual speech event: [k^{h}æ$^{?}$ts] *cats*. At the *phonemic* level, only contrasting speech sounds are represented: /kæts/. At the *morphophonemic* level, every morpheme has a unique representation: //kæt-P//, where //P// is a morphophoneme that abstracts over the plural allomorphs /-z/, /-s/, /-əz/, /-ən/ (*oxen*), /-rən/ (*children*), /-iː-/ (*geese*), etc.

In the theory of generative phonology (Chomsky and Halle 1968 – hereafter *SPE*), the *surface* level has approximately the same properties as the structuralists' allophonic level (though see Kingston (Ch.17) for discussion of some of the difficulties in pinning down the properties of the surface level). Generative phonology differs from structuralism, however, in denying that there are separate phonemic and morphophonemic levels, since positing this distinction leads to missed generalizations (Anderson 1985, Halle 1959). At generative phonology's *underlying* level, every morpheme

has a unique representation, except for suppletion. Underlying representations are composed of the same elements as surface representations, bundles of distinctive features, rather than phonetically uninterpretable symbols like the morphophoneme //P//. The English regular plural morpheme is /-z/, with suppletive alternants like /-ən/ or /-iː-/ listed lexically.

When a morpheme alternates non-suppletively, its underlying representation must be discovered by the analyst and 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 alternation: /bʊnt/ 'multi-colored' vs. /bʊnd/ 'federation'. In theory and in actual practice, as we will soon see, the relationship between the hypothesized underlying representation and the observed paradigm is sometimes less transparent than this.

Some recent research explores alternatives to posting 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 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 German, for example, final [t] in one paradigm member is allowed to correspond with non-final [d] in another member.

In this context, it is worth reviewing why generative phonology posits an underlying level of representation (see Kenstowicz and Kisseberth 1979: Ch.6 for an accessible 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 (1) is a well-known case.

(1) *Palauan Vowel Reduction*

Underlying	Present Middle	Future Participle (conservative)	Future Participle (innovative)	gloss
/daŋob/	[mə-ˈdaŋəb]	[dəˈŋob-l]	[dəŋəˈb-all]	'cover opening'
/teʔib/	[mə-teˈʔəb]	[təˈʔib-l]	[təʔəˈb-all]	'pull out'

Because unstressed vowels reduce to [ə] and there is only one stress per word, disyllabic roots like 'cover' and 'pull out' never show up with more than one surface non-schwa vowel. The hypothesized underlying representations /daŋob/ and /teʔib/ record the quality of the vowels as they appear when stressed. 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 those words to be derived by the grammar of the language.

In discussing the number and types of levels of representation that different theories allow, it is useful to introduce a distinction between what might be called designated and nondesignated levels. The designated levels are landmarks in a phonological derivation with special restrictions on their content or unique roles to play, particularly as the interface to other grammatical components. The nondesignated levels are usually not thought of as levels of representation at all; they are unremarkable points in the derivation lying intermediate between the designated levels.

Generative phonology in the *SPE* tradition recognizes only two designated levels of representation, underlying form and surface form, but it allows for any number of nondesignated levels intermediate between the underlying and surface levels. These nondesignated levels are the result of sequential application of phonological rules. *SPE* requires that all phonological rules apply sequentially. Therefore, 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, there is an intermediate level at which stress has been assigned but vowel reduction has not yet applied: /daŋob-l/ → daŋ́obl → [dəŋ́obl]. Indeed, *SPE* requires rules to apply sequentially even when simultaneous application would produce the same result (an exception is made for certain rules that can be conflated using *SPE*'s abbreviatory devices, which then must apply simultaneously). *SPE*'s intermediate levels do not have any special or unique roles, however; they are simply a side-effect of the way that rules apply, and so they will be referred to as nondesignated.

The theory of Lexical Phonology is firmly situated in the *SPE* tradition of rule application, but it imposes more structure on the grammar and increases the number of designated levels of representation (Kaisse and Hargus 1993, Kaisse and Shaw 1985, Kiparsky 1985, Mohanan 1982, among many others). In Lexical Phonology, the phonological grammar is organized, at a minimum, into separate lexical and postlexical modules, called strata. The output of the postlexical stratum is the surface representation, but the output of the lexical stratum is a designated intermediate level of representation with its own special properties. One of these properties, for example, is *structure preservation*, the requirement that the segments and structures occurring at this level be the same as those that are allowed in underlying representation. Depending on the language and on the specific version of Lexical Phonology applied to it, there may also be additional designated intermediate levels, such as a word-level stratum lying between the lexical and postlexical strata.

The theory of Lexical Phonology inherits from *SPE* the idea of sequential rule application and the resulting nondesignated levels of representation.

Alternative theories have been developed, however, in which sequential rule application is discarded but Lexical Phonology's modular structure is retained. These systems typically recognize just three levels, underlying, lexical or word, and surface. Approaches of this type include Harmonic Phonology (Goldsmith 1993a), Cognitive Phonology (Lakoff 1993), and Stratal Optimality Theory (5.4).

Apart from monostratal theories, the minimum number of levels of representation is of course two, underlying and surface. Finite-state phonological models, including a finite-state reduction of *SPE*, have this two-level property (Kaplan and Kay 1994, Karttunen 1993). More importantly for present purposes, Optimality Theory, as it was originally proposed by Prince and Smolensky (2004), maps underlying representations to surface representations with no intermediate levels.

5.3 Derivations

With the exception of monostratal theories, all current phonological models assume that the grammar maps underlying representations to surface representations. This mapping is called a *derivation*. Theories differ significantly in how complex derivations can be and in how derivations are organized internally.

The *SPE* approach to derivations retains considerable currency because it is often assumed even in contemporary theories that have moved far beyond *SPE*'s original hypotheses about rules and representations (e.g. Hayes 1995). 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 $i+1$. As was noted in Section 5.2, the outputs of individual rules constitute nondesignated levels of representation intermediate between underlying and surface form. The sole exception to this strict sequentiality is cyclic rule application, in which certain rules are allowed to reapply to successively larger grammatical constituents. (More will be said about cyclicity in Section 5.5.)

In *SPE*, the ordering of rules is *extrinsic*, which means that it is imposed on the rules by the grammar and cannot be predicted from rule form or function. From about 1969 through 1980, a voluminous literature developed around the question of whether some or even all aspects of rule ordering could be predicted (see Iverson 1995 for a brief survey or Anderson 1974 and Kenstowicz and Kisseberth 1977:chs.4,6 for more extensive discussion). A particular focus of attention in this period was the functional relationship between pairs of interacting rules: does one rule *feed* or *bleed* the other (Kiparsky 1968, 1976)?

Rule A is said to feed rule B if A creates 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, to be discussed in Section 5.4). 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 (2) shows, vowel epenthesis before a word-initial cluster (=rule A) creates new inputs to [ʔ] epenthesis (= rule B) before syllable-initial vowels.

(2) *Feeding order in Classical Arabic*

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

In the *SPE* model, the phonological grammar of Classic Arabic must include a statement to the effect that vowel epenthesis precedes [ʔ] epenthesis. In some revisions of that model (such as Anderson 1974, Koutsoudas, Sanders, and Noll 1974), this ordering statement was regarded as superfluous on the grounds that feeding order is unmarked or natural. In what sense is feeding order natural? If rules are allowed to apply freely at any point in the derivation when their structural descriptions are met, then the result will be the same as (2). Feeding orders maximize rule applicability. They also help to ensure that rules enforce generalizations that are surface-true: in Arabic, no syllable starts with a vowel because [ʔ] epenthesis applies freely.

Rule A is said to bleed rule B if A eliminates 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, also to be discussed in section 5.4). For example, in a southern Palestinian variety of Arabic, progressive assimilation of pharyngealization is blocked by high front segments, among them [i]. When the vowel [i] is epenthesized into triconsonantal clusters, it also blocks assimilation, as shown in (3a) (Davis 1995). Example (3b) is provided for comparison, since it shows progressive assimilation applying when it is unimpeded by intervening [i].

(3) *Bleeding order in southern Palestinian Arabic*

Underlying	(a)	/batˤnha/ 'her stomach'	(b)	/batˤn-ak/ 'your (m.sg.) stomach'
Vowel epenthesis		batˤinha		–
Progressive assimilation		–		batˤnˤaˤkˤ
Regressive assimilation		bˤaˤtˤinha		bˤaˤtˤnˤaˤkˤ
Surface		[bˤaˤtˤinha]		[bˤaˤtˤnˤaˤkˤ]

This is a bleeding order: epenthesis eliminates some opportunities for progressive assimilation to apply. In the *SPE* model, the phonological grammar of Palestinian Arabic must include a statement to the effect that vowel

epenthesis precedes progressive assimilation. Bleeding orders do not maximize rule applicability: on the contrary, the bleeding order in (3a) robs progressive assimilation of a chance to apply. But bleeding orders do help to ensure that rules state surface-true generalizations: the effect of the bleeding order in (3a) is that progressive assimilation does not traverse any surface [i] vowel, regardless of whether it is present in the input or derived by rule.

As these remarks suggest, feeding and bleeding interactions have something in common: when feeding and bleeding orders are in effect, structures derived by a rule are treated exactly the same as structures that were already present in underlying representation. For example, the derived initial vowel in the intermediate representation [id^{ˤ}rib] is treated the same as the underlying initial vowel in /al-walad-u/ 'the boy (nom.sg.)'; both trigger [ʔ] epenthesis, yielding [ʔid^{ˤ}rib] and [ʔalwaladu]. Likewise, epenthetic and non-epenthetic [i] equally block progressive assimilation in Palestinian, as shown by (3a) and /s^{ˤ}iħħa/ → [s^{ˤ}iħħa] '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, which will be discussed in Section 5.4.

Because simple feeding and bleeding interactions yield surface-true generalizations, the intermediate derivational stage is superfluous. Therefore, examples like (2) and (3) can be readily accommodated in theories that posit much shallower derivations than the *SPE* model. Although the discussion here will focus on Optimality Theory, much the same can be said about any of the other approaches mentioned at the end of Section 5.2.

The central idea of OT is that constraints on linguistic forms are ranked and violable. Constraints come in two types: markedness constraints impose restrictions on surface representations, and faithfulness constraints require identity in the mapping from underlying to surface form. In feeding-type interactions, two markedness constraints are active, with both dominating antagonistic faithfulness constraints. In the Classical Arabic example (2), the active markedness constraints are *Complex, which prohibits tautosyllabic clusters, and Onset, which prohibits vowel-initial syllables. Both dominate the faithfulness constraint Dep, which militates against epenthesis. The ranking argument is given in (4).

(4) **Complex, Onset Dep*

/dˤrib/	*COMPLEX	ONSET	DEP
☞ (a) ʔidˤrib			* *
(b) idˤrib		*!	*
(c) dˤrib	*!		

Because satisfying *Complex creates a condition that puts Onset in peril, as shown by candidate (4b), there is no need to go through an intermediate step where vowel epenthesis has occurred but consonant epenthesis has not. It is enough to say that surface forms must satisfy both of these constraints, even at the expense of unfaithfulness to the input.

When two rules contradict one another, at least in part, their relationship does not fit the simple feeding/bleeding classification. An example comes from Nuuchahnulth, formerly known as Nootka (Sapir and Swadesh 1978). This language has a process that rounds velars and uvulars when they follow round vowels (5a), as well as a process that unrounds velars and uvulars at the end of a syllable (5b). When a velar or uvular consonant is preceded by a round vowel and also falls at the end of a syllable, these two rules are in conflict, a conflict that the *SPE* model resolves by ordering them as in (5c). The result is that consonants surface as nonround when they both follow a round vowel and precede a syllable boundary (indicated by a period/full stop).

(5) *Nuuchahnulth (un)rounding*

(a)	Rounding		
	Underlying	/ħaju-qi/	'ten on top'
	Rounding	[ħa.ju.q^{w}i]	(cf. [hi.ta.qi] 'on top')
(b)	Unrounding		
	Underlying	/ɬakː^{w}-ʃit͡ɬ/	'to take pity on'
	Unrounding	[ɬaː.k.ʃit͡ɬ]	(cf. [ɬaː.k^{w}iq.nak] 'pitiful')
(c)	Interaction		
	Underlying	/m'uːq/	'throwing off sparks'
	Rounding	[m'uːqw]	(cf. [m'o.q^{W}ak] 'phosphorescent')
	Unrounding	[m'uːq]	

Pullum (1976) dubs this a *Duke-of-York derivation*, after the English nobleman who, in a nursery rhyme, orders his men up a hill and then down again (also see Kenstowicz and Kisseberth 1977:171ff.) These rules are in a mutual feeding relationship, and it is not possible for both of them to state surface-true generalizations. Under *SPE* assumptions, the 'truer' rule is the one that is ordered last, syllable-final unrounding.

In OT, because constraints are ranked and violable, there is no need to go through an intermediate stage where the consonants become rounded, only to lose that rounding later in the derivation. The Nuuchahnulth situation involves conflict between two markedness constraints, one requiring that velars and uvulars be nonround at the end of a syllable (call it *K^{w}]$_{\sigma}$), and the other requiring that they be round after a round vowel (call it *uK). Faithfulness to rounding is ranked below both of these markedness constraints. The ranking argument is shown in (6).

(6) $^{*}K^{w}]\sigma \gg {}^{*}uK \gg$ IDENT(ROUND)

/m'u:q/	$^{*}K^{w}\ ]_{\sigma}$	*uK	IDENT(round)
☞ (a) m'u:q		*	
(b) m'u:q^{w}	*!		*

The Nuuchahnulth example further illustrates why OT, in its original conception, maps underlying representations directly to surface representations, without intermediate levels. In the *SPE* model, ordering is a way of establishing priority relationships among rules, and in a case like Nuuchahnulth it is the last rule that has priority in the sense that it states a surface-true generalization, even though the earlier rule does not. In OT, priority relationships among constraints are established by ranking them, and (6) shows that ranking can replace at least some applications of rule ordering. The strongest claim, then, is that OT can dispense with ordering and all of its trappings, including intermediate derivational steps. This claim is not uncontroversial (see Section 5.4).

The discussion in this section suggests that sequential rule application is unnecessary, at least for feeding and bleeding interactions. The evidence of counterfeeding and counterbleeding interactions will be discussed in the Section 5.4, but first it is necessary to remark on certain conceptual arguments that have been made in support of sequential rule application.

One of these conceptual arguments holds that sequential rules accurately model a system of mental computation (Bromberger and Halle 1997). The failure of the Derivational Theory of Complexity showed that this idea is very far off the mark, at least in syntax (Fodor, Bever, and Garrett 1974); the same seems to be true in phonology (Goldsmith 1993b). Indeed, if the goal of generative grammar is to construct competence models (Chomsky 1965), then it is a category mistake to ask whether these models faithfully replicate mental computation.

Another argument offered in favor of sequential rule application is that it makes sense in terms of language history (Bromberger and Halle 1989): the ordering of synchronic rules matches the chronology of diachronic sound changes. The problem with this view is that it somewhat misconceives the diachronic situation. If generation X+1 innovates a sound change, they do not simply add a rule onto the end of generation X's phonological grammar – they cannot, since generation X+1 obviously does not have direct access to generation X's grammar. In other words, generation X+1's learning is informed exclusively by X's productions, as filtered through the X+1 perceptual system. X's productions offer only indirect evidence of X's grammar, subject to well-known limitations like the absence of positive evidence. From this perspective, we neither expect nor do we necessarily observe that grammars change by accreting rules at the end of the ordering.

5.4 Opacity

If rule A feeds rule B but they are applied in the order B precedes A, then these rules are said to be in *counterfeeding* order. For example, in a variety of Bedouin Arabic (Al-Mozainy 1981, McCarthy in preparation), 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 (7).

(7) *Counterfeeding order in Bedouin Arabic*

Underlying	(a)	/dafaʕ/ 'he pushed'	(b)	/ʃarib-at/ 'she drank'
Deletion		–		ʃarbat
Raising		difaʕ		–
Surface		difaʕ		ʃarbat

High vowels derived by raising (7a) are treated differently from underlying high vowels (7b); only the underlying high vowels are subject to deletion. In a feeding order like (2), derived and underlying structures behave alike, but in a counterfeeding order they behave differently.

The same is true of *counterbleeding* order, where rule A bleeds rule B but they are applied in the order B precedes A. In this same Arabic dialect, there is also a process palatalizing velars when they are adjacent to front vowels. 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 (8).

(8) *Counterbleeding order in Bedouin Arabic*

Underlying	(a)	/ħaːkim-iːn/	(b)	/t-ħa–kum-in/
Palatalization		ħaːkʲimiːn		–
Deletion		ħaːkʲmiːn		t-ħakm-in
		'ruling (m.pl)'		'they (f.) rule'
		(cf. [ħaːkʲim]		(cf. [t-ħakum]
		'ruling (m.sg.)')		'you (m.sg.) rule')

High front vowels, even when they are absent from surface forms, induce adjoining velars to palatalize. Example (8b) shows the necessary contrast: a velar is not palatalized in a virtually identical surface context that is derived from a different underlying source with a back rather than a front vowel.

The result of counterfeeding and counterbleeding interactions is phonological *opacity*. Kiparsky (1976) defines opacity as in (9).

(9) *Opacity*

A phonological rule P of the form $A \rightarrow B \mid C\_\_D$ 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*.
(c) instances of *B* not derived by P that occur in the environment *C*__*D*.

Clause (9c) describes all processes of neutralization and so it is not relevant to our concerns here. We will focus then on clauses (9a) and (9b).

In the derivation /dafaʕ/ → [difaʕ] (7a), the deletion rule is opaque under clause (9a) of this definition: there are instances if [i] (=*A*) in an open syllable (=*C*__*D*). Typically, counterfeeding order produces opacity of this type, in which surface forms contain phonological structures that look like they should have undergone some rule but in fact did not.

In the derivation /ħaːkimiːn/ → [ħaːkʲmiːn] (8a), the palatalization rule is opaque under clause (9b) of this definition: there are instances of [kʲ] (=*B*) derived by palatalization that are not in this rule's context, adjacent to a front vowel (=*C*__*D*). Typically, counterbleeding order produces opacity of this type, in which surface forms contain derived phonological structures without the context necessary for them to be derived.

Counterfeeding and counterbleeding interactions supply the best (arguably, the only) evidence for language-particular rule ordering. It is not surprising, then, skepticism about stipulated rule ordering stimulated efforts to deny that opaque interactions involve living phonological processes. According to the proponents of Natural Generative Phonology (Hooper [Bybee] 1976, 1979, Vennemann 1974), real phonological rules must state surface-true generalizations and they must be unordered. They therefore maintain that opaque processes are merely the lexicalized residue of sound changes that are no longer productive– the commonly-used phrase is that they are not "psychologically real". In fact, much if not all of the abstractness controversy of the 1970s, which dealt with proposed limits on the degree of disparity between underlying and surface representations (see Kenstowicz and Kisseberth 1977:Ch.1, 1979:Ch.6), was really an argument about opacity, since underlying forms are abstract precisely because opaque rules operate on them.

Certainly, there have been dubious analyses based on opaque rules and excessively abstract underlying forms (*SPE*'s /rixt/ → [raːjt] *right* comes to mind – Chomsky and Halle 1968:233–4), but complete denial of opaque interactions is an overreaction. The Bedouin Arabic example is instructive. 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 applies productively in across word boundaries (10), which means that it cannot be lexicalized.

(10) *Phrase-level deletion in Bedouin Arabic*

/ʃaːrib al-maː/	[ʃaːrb alma]	'drinking the water'
/tiʕtˤuːnih al-musayʕiːdiː/	[tiʕtˤuːnh almseːʕiːdi]	'you give it to the one from the clan of Musa' īd'

Third, the most compelling evidence that raising is productive comes from a secret or play language. Although raising generally affects short /a/ in a non-final open syllable, there are phonological conditions under which it regularly fails to apply: after a guttural consonant ([h], [ħ], [ʕ], [ξ], [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 alter the conditions necessary for raising. When this happens, the vowel raises or not in exact conformity with these generalizations (11):

(11) *Raising alternations in a secret language*

difaʕ	fidaʕ	
	daʕaF	No raising before guttural + [a]
	faʕad	"
	ʕadaf	No raising after guttural
	ʕafad	"

Fourth, the secret language data show that palatalization is also productive, even though it is opaque. In sum, the opaque phonology of Bedouin Arabic is also its living phonology. (For further examples of processes that are productive yet opaque, see Donegan and Stampe 1979.)

If opacity is an authentic property of phonology, then any successful phonological theory must be able to accommodate it, at least in robust instantiations like Bedouin Arabic. Theories of the *SPE* variety, with as many levels of representation as there are rules, have no difficulty with opacity, as we have seen. The challenge, then, is to account for opacity within theories whose resources are more limited. There is certainly no consensus about how best to do this, but there are several promising lines of on-going research.

The most direct line of attack on the opacity problem is to retain something like the basic rule-ordering mechanism but limit the theory to three or four designated levels of representation, with no nondesignated levels. For example, Harmonic Phonology (Goldsmith 1993a) and Cognitive Phonology (Lakoff 1993) recognize just three levels of representation, called morphophonemic (M), word (W), and phonetic (P). The M and P levels are equivalent to underlying and surface representation, respectively; the innovation is to recognize a unique intermediate level, W. Processes that occur in the M→W mapping necessarily precede processes that occur in the W→P mapping, so limited effects of rule ordering can be achieved.

Stratal Optimality Theory obtains opaque interactions similarly (Kiparsky 2000, 2003, McCarthy and Prince 1993b, Rubach 2000, and contributions to

Hermans and Oostendorp 1999 and Roca 1997a, among many others). Stratal OT is also called OT/LP because of its connection with the rule-based theory of Lexical Phonology. The basic idea is that a succession of OT grammars is linked serially, with the output of one grammar constituting the input to the next one. These grammars are distinct, which in OT means that they contain different rankings of the same universal constraint set. Each of these grammars corresponds to one of the strata of Lexical Phonology; this includes one or more lexical strata, a word stratum, and a postlexical stratum, which altogether define at least four levels of representation. As in Harmonic and Cognitive Phonology, opaque interactions are obtained by the intrinsic ordering between these grammar modules.

The counterbleeding interaction of palatalization and deletion in (8) will serve to illustrate Stratal OT in action. This interaction requires that the /k/ → [k^j] unfaithful mapping occurs in a stratum earlier than the /i/ → Ø unfaithful mapping. If the /k/ → [k^j] mapping is the result of a ranking that holds in the word stratum, then the constraint ranking responsible for deletion must not obtain until the postlexical stratum. This system is illustrated with the tableaux in (12). In these tableaux, eletion of high vowels is assumed to be a response to the markedness of high vowel nuclei under *NUC/[HI], following Gouskova (2003); velar palatalization is attributed to the cover constraint PAL, which prohibits sequences of a plain velar and a front vowel.

(12) *Stratal OT approach to opacity in Bedouin Arabic*

(a) Word stratum: PAL, MAX » *K^j » IDENT(back); MAX » *NUC/[HI]

/ħaːkim-iːn/	PAL	MAX	*K^j	*NUC/[HI]	IDENT(back)
☞ (a) ħaːk^{j}imiːn			*	* *	*
(b) ħaːkimiːn	*!			* *	
(c) ħaːkmiːn		*!		*	
(d) ħaːk^{j}miːn		*!	*	*	*

(b) Postlexical stratum: IDENT(back) » *K^j; *NUC/[HI] » MAX

/ħaːk^{j}imiːn/	IDENT(back)	PAL	*NUC/[HI]	*K^j	MAX
(a) ħaːk^{j}imiːn			**!	*	
(b) ħaːkimiːn	*!	*!	**!		
(c) ħaːkmiːn	*!		*		*
☞ (d) ħaːk^{j}miːn			*	*	*

The word stratum (12a) requires the ranking PAL » *K^j » IDENT(back), which is necessary to explain why palatalized velars occur only in contiguity with (underlying) front vowels. It also requires the ranking MAX » *NUC/[HI]. This ranking prevents deletion in the word stratum, since if deletion were allowed then the transparent form *[ħaːkmiːn] would win. In the postlexical stratum (12b), two rerankings are necessary. The ranking of *NUC/[HI] and MAX must be reversed so that deletion takes place in the postlexical

phonology. The other reranking, that of IDENT(back) and *K^{j}, is necessary to prevent depalatalization of the previously palatalized velar. Since the input to the postlexical stratum is the output of the word stratum, IDENT(back) is protective of the derived [k^{j}] in the word-stratum output / postlexical-stratum input [ħaːk^{j}imiːn].

Stratal OT's approach to opacity is a significant departure from the original theory of Lexical Phonology, which recognized two possible sources of opaque ordering: the intrinsic ordering of rules that are assigned to different strata, and the extrinsic ordering of rules within a stratum. Stratal OT makes a much stronger claim: all opaque interactions are reducible to processes that occur transparently in different strata. It remains to be seen whether this claim survives empirical scrutiny, including the challenge presented by extant Lexical Phonology analyses that require within-stratum opaque ordering, such as Kiparsky's (1984) analysis of Icelandic (also see Noyer 1997:515, Paradis 1997:542, Roca 1997b:14ff., Rubach 1997:578 for various critical remarks).

Stratal OT and rule-based Lexical Phonology agree on a different claim: if independent criteria require that two processes be assigned to different strata, then the ordering of those processes is forced by the intrinsic ordering of the strata. In Lexical Phonology, there were many criteria that tended to segregate processes by stratum, such as structure preservation or the strict cycle. Stratal OT has abandoned nearly all of these principles, but one remains: the stratum determines the domain of a process. Processes that can apply between words are necessarily postlexical, whereas processes that are word-bounded are necessarily assigned to the lexical or word strata. The counterfeeding interaction in (7) presents a direct challenge to this claim. Raising is word-bounded; except for a few fixed expressions like /baːrak al^{ˤ}l^{ˤ}ah fiːk/ → [baːr^{ˤ}ik al^{ˤ}l^{ˤ}ah fiːk] 'may Allah bless you', raising does not occur across word boundaries even when an open syllable is created by syllabifying a word-final consonant as an onset when the next word begins with a vowel. Deletion is a phrase-level process (10), so it must be assigned to the postlexical stratum, as we have already noted. Since the word stratum where raising occurs precedes the postlexical stratum where deletion occurs, raising should feed deletion, resulting in derivations like /samiʕ-t/ $\rightarrow_{\text{Word}}$ [simiʕt] $\rightarrow_{\text{Postlex}}$ *[smiʕt] 'you (m.sg.) heard'. The correct form is [simiʕt], since raising does not in fact feed deletion. Furthermore, there is no straightforward way to salvage the analysis, since the failure is one of principle. For deletion to be in a counterfeeding relationship with any other process, that process must be assigned to a stratum later than deletion's stratum, but since deletion is a phrase-level process, there is no later stratum. It would seem, then, that no analysis is possible within the assumptions of Stratal OT.

Targeted constraints (Wilson 2000), comparative markedness (McCarthy 2002a, 2003a), sympathy (McCarthy 1999, 2003b), and virtual phonology

(Bye 2001) also rely on a third representation, neither underlying nor surface, to support the analysis of opacity in OT. These various approaches differ from each other and from stratal OT in how they organize the grammar and how they identify that third form, but at a sufficiently distant level of abstraction they share this point of similarity.

Space does not permit a thorough review of these approaches, their advantages, and their limitations, so a brief sketch will have to suffice, using as an example the counterfeeding interaction of raising and epenthesis in Bedouin Arabic. Raising occurs in open syllables, but open syllables derived by epenthesis do not condition raising: /gabr/ → [gabur^{ʕ}], *[gibur^{ʕ}] 'a grave'. The third form that indirectly influences the outcome is *[gabr^{ʕ}], an output representation that lacks the epenthetic vowel. Targeted constraints are inherently comparative, and *[gabr^{ʕ}] is the basis for comparison by a constraint that says, in effect, that a word without a final cluster is more harmonic than an otherwise identical word with that cluster, so [gabur^{ʕ}] ≻ *[gabr^{ʕ}]. In comparative markedness, the constraint responsible for raising asks whether [a] is in an open syllable in the fully faithful candidate *[gabr^{ʕ}]. Sympathy theory looks to the candidate that is most harmonic except that it obeys DEP, and this too is *[gabr^{ʕ}]. Virtual phonology selects *[gabr^{ʕ}] as the third or 'virtual' form using markedness and faithfulness constraints that are indexed to the virtual evaluation. In short, these various theories share the assumption that the form *[gabr^{ʕ}], *qua* output, exerts indirect influence over the outcome of harmonic evaluation, so that opaque [gabur^{ʕ}] triumphs over transparent *[gibur^{ʕ}]. (For critical discussion of targeted constraints, see McCarthy (2002b); of comparative markedness, see the various rejoinders appearing in *Theoretical Linguistics* 29 (2003); of sympathy, see Itô and Mester (2001), Kiparsky (2000), and McMahon (2000a).)

Another general strategy for attacking the opacity problem is to allow rules or constraints to have simultaneous access to different levels of representation. A classic *SPE* phonological rule has an elementary form of this property: its structural description is met at some (nondesignated) level of representation, and its structural change creates the next level of representation after that. Variations on this scheme can accommodate differences between transparent and opaque interactions. For example, Harmonic and Cognitive Phonology provide a system of two-level rules (also see Karttunen 1993, Koskenniemi 1983). A two-level rule can specify a structural description that must be met by its input, its output, or both. In Bedouin Arabic, for example, the structural description of raising requires that the affected vowel be in an open syllable in the input (13a), since open syllables derived by vowel epenthesis do not condition raising: /gabr/ → [gabur^{ʕ}] 'a grave'. On the other hand, the transparent interaction of vowel and consonant epenthesis in (2) shows that the structural description of consonant epenthesis must be met in the output (13b).

(13) *Some two-level rules*

(a) $a]_{\sigma}$ ↓ i

(b) Ø ↓ $_{\sigma}[$ʔV

By their very nature, faithfulness constraints in OT have access to two levels of representation, and so it is not surprising that extensions of the basic faithfulness theory have been applied to opacity. Constraint conjunction is a mechanism for combining constraints: the constraint $C = [A \& B]_{\delta}$ is violated if and only if some constituent or sequence of type δ violates both A and B. The conjunction of two faithfulness constraints produces a type of faithfulness constraint that can be applied to counterfeeding opacity. For example, $[\textsc{Ident}(\text{low}) \& \textsc{Dep}]_{\textsc{Adj-}\sigma}$ is violated if a vowel is raised and a vowel is epenthesized in adjacent syllables. Ranked appropriately, this constraint will rule out the mapping /gabr/ → *[gibur^{ˤ}] while still allowing /dafaʕ/ → [difaʕ], where there is no nearby epenthesis. The problem with local conjunction is that it rules out the cooccurrence of unfaithful mappings in close proximity, but mere proximity is not the source of opacity. Rather, counterbleeding opacity involves unfaithful mappings that crucially *interact* with one another; what is forbidden is for epenthesis to create the open syllable that conditions raising. The difference between proximity, which has no apparent linguistic relevance, and interaction, which is the basis for opacity, becomes clear once it is realized $[\textsc{Ident}(\text{low}) \& \textsc{Dep}]_{\textsc{Adj-}\sigma}$ is violated not only in the interacting case *[gibur^{ˤ}], where epenthesis creates the open-syllable context for raising, but also when epenthesis occurs in the preceding syllable, where it does not interact with raising. This prediction of the local-conjunction model is not only typologically implausible – in known cases of counterfeeding opacity, interaction and not proximity is essential – but also factually incorrect in Bedouin Arabic, as shown by examples like /t^{ˤ}arad ʁanam-i/ → [t^{ˤ}aradi̱ʁni̱m-i] 'I pursued my sheep'. Here, the first underlined [i] is epenthetic and the second is the result of raising, showing that there is no prohibition on raising a vowel when there is epenthesis in the preceding syllable.

Another way of allowing simultaneous access to two levels of representation is to fold them into a single level of representation (for a monostratal approach to opacity within Declarative Phonology, see Bye 2003). The development of nonlinear phonology in the 1970s offered ways of making distinctions between underlying and derived structures that would otherwise be identical, and Prince and Smolensky's (2004) Parse/Fill model of faithfulness exploits this possibility. One assumption of this model is that segments are never literally deleted; rather, they remain present in the segmental string but are unpronounced because they are not incorporated into prosodic structure. The lingering presence of the underlying but unprononounced segment offers opportunities for the transparent analysis of opaque interactions. In the Bedouin Arabic counterbleeding case (8), for

instance, the winning candidate has an unsyllabified [<i>] that transparently induces palatalization of the preceding velar: [ħaːkʲ<i>miːn]. (For further developments along these general lines, see Goldrick (2000).)

Finally, it is worth noting that opaque interactions contribute in a backhanded way to maintaining the transparency of the input-output relation. For example, the speaker of Bedouin Arabic who hears [gaburˤ] can legitimately infer that the [u] is epenthetic, since that is why the preceding [a] is not raised. Kaye (1974, 1975) and Kisseberth (1973) discussed such functional motivations for opacity, and Lubowicz (2003) has developed an OT-based system in which opacity serves to preserve underlying contrasts.

This review of opacity does not exhaust a very rich topic, and future developments can surely be expected. There is a need for a body of solidly supported examples of phonological opacity, similar to Bedouin Arabic, and for greater understanding of the nature of and limits on opaque interaction.

5.5 Cyclicity

In *SPE*, the strict linear order of phonological rules admits of a single exception: cyclic rule application. Certain rules are 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 those in (14), From American English:

(14) *Transderivational similarities*

(i) Monomorphemic words like *ˌKalamaʼzoo* and *ˌWinnepeʼsaukee* show the normal stress pattern when three light syllables precede the main stress. Derived words like *acˌcrediʼtation* and *iˌmagiʼnation* deviate from this pattern under the influence of *acʼcredit* and *i ʼmagine*.

(ii) A closed, sonorant-final syllable is normally unstressed in pre-stress position: *ˌserenʼdipity*, *ˌgorgonʼzola*, *ˌPennsylʼvania*. But the same kind of syllable is stressed in the derived words *ˌauˌthenʼticity* and *ˌconˌdemʼnation* under the influence of *ˌauʼthentic* and *conʼdemn*.

In *SPE*, the aberrant stress of the 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 a new primary stress is assigned later in the word. Monomorphemic *Kalamazoo* and *serendipity* have no inner cycle, so they show the effects of just a single pass through the stress rules.

Cyclic rule application has also been invoked to account for prosodic closure effects that have no obvious transderivational motivation. In Axininca Campa, for example, /V+V/ sequences at stem+suffix juncture are syllabified by epenthesizing [t] (Payne 1981): /i-N-koma-i/ → [iŋ.ko.ma.t̪i] 'he will paddle'; /i-N-koma-ako-i/ → [[iŋ.ko.ma.t̪a.ko.t̪i] 'he will paddle for'. Since *[iŋ.ko.mai] and *[iŋ.ko.maː.koi] are phonotactically possible in this language, the problem comes down to explaining why a syllable like [mai] is forbidden just in case [ma] and [i] come from different morphemes. Spring (1990) proposes an analysis based on cyclic syllabification: the stem [iŋ.ko.ma] is fully syllabified on the inner cycle, and on the outer cycle affixal [i] is by assumption barred from joining any pre-existing syllable, forcing it to join with epenthetic [t] to become syllabified. Cyclic syllabification explains why vowel-final stems are closed under syllabification. Because Axininca Campa does not allow final codas, consonant-final stems cannot be closed under syllabification. Instead, the final consonant remains extrasyllabic until affixal /-i/ is added on the next cycle, at which point they join to form a syllable: /i-N-t͡ʃʰ ik-i/ $\rightarrow_{\text{1st cyc.}}$ [iŋ.t͡ʃʰi.<k>] $\rightarrow_{\text{2nd cyc.}}$ [iŋ.t͡ʃʰi.ki]. Hence, consonant-final stems are not prosodically closed.

Cyclic effects of both types have attracted a great deal of recent attention, particularly in OT. Three basic approaches can be identified and will be discussed in turn. It should be noted that these approaches are not necessarily inconsistent with one another; they may be complementary, each with its own proper analytic domain.

Closest to the *SPE* model are those accounts that regard transderivational relationships as fundamentally asymmetrical: if word or stem A exerts an influence on the phonology of word or stem B, then B cannot exert an influence on A. Typically, A and B stand to one another as base and derivative, like *authentic* and *authenticity*. This can be accomplished by combining an *SPE*-style cycle with an OT grammar, taking the output of the grammar, adding an affix, and then returning the result to the grammar as a new input. It can also be done with *output-output faithfulness* constraints, which require that related words resemble one another, just as ordinary faithfulness constraints demand identity between input and output (Benua 1997, Kager 1999b, Pater 2000b). A strength of output-output faithfulness is its restrictiveness, limiting cyclic effects to transderivational relationships between actually existing words. A weakness is the need to stipulate the asymmetry with a principle of 'base priority'.

More distant from *SPE* and stratal OT are approaches that allow symmetric transderivational effects: word B can also influence the phonology of word A even if, morphologically, B is derived from A. Burzio (1994) and Kenstowicz (1996) were early advocates of this view; Downing, Hall, and Raffelsiefen (2005) is a recent anthology containing much relevant work. Symmetric transderivational effects seem to be important in inflectional paradigms. Morphologically, paradigms lack the obvious base/derivative structure of derivational morphology. In the Classical Arabic perfective

verb paradigm (15), for example, there is little reason to see one form as more basic than the others:

(15) *Classical Arabic perfective paradigm of ktb 'write'*

	singular	plural
1st	katabtu	katabnaː
2nd masc.	katabta	katabtum
2nd fem.	katabti	katabtunna
3rd masc.	kataba	katabuː
3rd fem.	katabat	katabna

The transderivational effect exhibited by the Arabic paradigm involves the impossibility of having a verb stem with a long vowel in the second syllable (McCarthy 2005). Some members of the paradigm have suffixes that begin with consonants, such as [katabtu], and other members have suffixes that begin with vowels, such as [kataba]. If it were possible to have a verb stem with a long vowel in the second syllable, then its paradigm would necessarily have a vowel length alternation, because long vowels are shortened in closed syllables: the paradigm for the hypothetical stem /tabaːk/ would include [tabaktu], [tabaːka], etc. But there are no such verbal paradigms in the language, indicating that some constraint rules out vowel length alternations within paradigms. In other words, [tabaːka] is ill-formed because it differs in vowel length from its paradigmatic relative [tabaktu], or more generally the stems with vowel-initial suffixes must accommodate themselves, as regards vowel length, to the stems with consonant-initial suffixes, where vowel length is excluded for phonological reasons. It is risible to suggest, as a strict commitment to asymmetry would demand, that some stem with a consonant-initial suffix is the base from which all other stems are derived. Rather, information about phonological form flows freely in any direction within a paradigm, even between forms with no obvious base/derivative relationship.

Finally, prosodic closure phenomena like the one in Axininca Campa are amenable to analysis using alignment constraints (McCarthy and Prince 1993a). Alignment constraints require that the edges of morphological and prosodic constituents coincide. One such constraint, ALIGN-R(stem, σ), says that the rightmost segment in every stem must be final in some syllable. In Axininca Campa, it crucially dominates DEP, so it is able to compel consonant epenthesis (16):

(16) *Align-R (stem, σ) ≫ DEP*

/i-N-koma-i/	ALIGN-R (Stem, σ)	DEP
☞ (a) iŋ.ko.ma|.ti		*
(b) iŋ.ko.ma|i	*!	

The failed candidate (16b) has an unaligned stem that ends in mid-syllable (the right edge of the stem is indicated by the vertical bar). The winner (16a) lines up the stem and syllable exactly at the expense of epenthesizing a consonant. Though decisive in /V+V/ junctures like this, ALIGN-R(stem, σ) is crucially dominated by a restriction on coda consonants, CODA-COND. That is why there is no consonant epenthesis in /C+V/ juncture (17):

(17) *CODA-COND ≫ ALIGN-R (stem, σ)*

/i-N-t͡ʃʰik-i/	CODA-COND	ALIGN-R (Stem, σ)	DEP
☞ (a) iŋ.t͡ʃʰi.k|i		*	
(b) iŋ.t͡ʃʰik|.ti	*!		*

Cyclic or transderivational relationships are one aspect of the larger topic of how phonology interfaces with the other grammatical components, morphology and syntax (see Ussishkin Ch.19, Urbanczyk Ch.20, Truckenbrodt Ch.18). Cyclicity also has connections with the opacity problem, connections that are made quite explicitly in stratal OT.

5.6 Conclusion

This chapter has explored the concept of level of representation and the closely related idea of a derivation that connects the different levels of representation with one another. These are areas of on-going, productive research activity. As this work continues, we may expect to see some consensus emerging about the basic questions: How many and what kind of levels of representation are there? Are serial derivations a central property of phonology, and if so what are their properties? What is the range and character of opacity phenomena, and how are they best analyzed? How do morphological structure and morphological relatedness impinge on phonology?