Phonology in Universal Grammar[1]

Brett Miller, Neil Myler, and Bert Vaux

April 2014

1. Introduction: definitions and scope

In order to investigate the phonological component of Universal Grammar (UG), we must first clarify what exactly the concept of UG involves. The terms "Universal Grammar" and "Language Acquisition Device" (LAD) are often treated as synonymous,[2] but we believe that it is important to distinguish between the two. We take a *grammar* to be a computational system that transduces conceptual-intentional representations into linear (but multi-dimensional) strings of symbols to be interpreted by the various physical systems employed to externalize linguistic messages. It thus includes the traditional syntactic, morphological, and phonological components, but not phonetics, which converts the categorical symbols output by the grammar into gradient representations implementable by the body.

Bearing the above definition of "grammar" in mind, we take "*Universal* Grammar" to refer specifically to the initial state of this computational system that all normal humans bring to the task of learning their first language (cf. Hale and Reiss (2008:2)). The phonological component of this initial state may contain, inter alia, rules (the "processes" of Natural Phonology (Stampe (1979))), violable constraints (as in Calabrese's (1988, 1995) marking statements or Optimality Theory (OT)'s markedness and faithfulness constraints (Prince and Smolensky (1993/2004))), a set of parameters (e.g. Dresher's (2009) feature tree), or nothing at all (Chomsky and Halle (1968), Reiss (2008); what these call Universal Grammar falls under our definition of the LAD to be set out below). Issues of interest in the phonological literature relevant to this conception of UG include (in OT) the contents of the universal constraint set CON, (in Natural Phonology) the set of natural processes such as coda devoicing, (in Calabrese's (1988/1995) model) the inventory and ranking of marking statements, and so on.

We take the Language Acquisition Device on the other hand to be the complex of components of the mind/brain involved in hypothesizing grammar+lexicon pairs upon exposure to primary linguistic data (PLD), the subset of sensory percepts deemed by learners to be valid tokens of the language they are attempting to learn. This conception of the LAD includes (limiting ourselves for present purposes to items of phonological interest) the mechanisms that determine what aspects of sensory input are linguistic, the system(s) that transduce sensory input into linguistic symbols (distinctive features, prosodic units, linear precedence encoders), and the inventories of symbols, relations, and operators which these systems manipulate. The latter three inventories are commonly assumed to include a universal set of distinctive features, a set- or tree-theoretic hierarchy of relations between these features (feature geometry, Clements (1985)), a universal set of logical operators such as ∀, ∃, and ¬ that can take scope over these elements (Reiss (2003), Fitzpatrick, Nevins, and Vaux (2004)), and (in Rule-Based Phonology) operators encoding the notions SPREAD, DELINK, and

[1] Authors are listed in alphabetical order. Thanks to James Clackson, Ahrong Lee, Dave Odden, Ian Roberts, Bridget Samuels, Oktor Skjærvø, Patrick Taylor, and Jeffrey Watumull for helpful comments on aspects of this chapter.

[2] Compare for instance Chomsky's (1968, 1972, 2006) characterization of the LAD with his 2007 definition of UG as "the theory of the initial state of FL [the faculty of language]", which encompasses the senses of both UG and the LAD defined below.

INSERT (cf. McCarthy (1988)) or (in OT) GEN (the function that generates a potentially infinite set of candidate outputs from a given input via an unspecified set of operations) and EVAL (the function that assigns violation marks to output candidates and selects winners).

Issues of interest in the phonological literature relevant to this conception of the LAD include learning biases that constrain the transparency (Kiparsky (1973)), granularity (Pierrehumbert (2001)), extension (Hale and Reiss (2008), Ross (2011)), and ordering (Koutsoudas, Sanders, and Noll (1974); Kiparsky (1985) inter alia) of phonological hypotheses. Also of interest are the inventories and internal organization of phonological symbols, the range of possibilities for combining these symbols, the ontology of recurrent constraints such as the Obligatory Contour Principle (Leben (1973)), Derived Environment Constraint (Kiparsky (1993)), Sonority Sequencing Principle (Clements (1990)), and Final Consonant Extraprosodicity (Itō (1986)), the mechanisms used to construct underlying representations (Vaux (2005), Nevins and Vaux (2008)), and so on.

Both UG and LAD understood in the senses just described present a wide range of questions worthy of phonological investigation and debate. Since the semantic domain of the term "UG" in the literature typically encompasses both UG and the LAD as defined above (e.g. White 2003:xi, 2), and since what counts as part of UG vs. the LAD differs depending on the linguistic architecture one assumes (e.g. the putative final voicing gap investigated in section 4.2 would be attributed by OT to mechanisms in UG, but in the LAD in classic Rule-Based Phonology (RBP; Kenstowicz 1994)) we touch on aspects of both in this chapter.

Research involving UG is typically cast in terms of whether or not such an entity exists (see e.g. Cowie (1999) and response by Fodor (2001)), but we consider this to be a non-question, as learners demonstrably must bring a non-trivial array of prior knowledge to the learning task if anything at all is to be learned (cf. Yang (2004:451))—to put the matter baldly, an actual blank slate learns nothing whatsoever upon exposure to PLD. In practice, the questions being debated between the nativists and empiricists typically have to do with the exact nature and extent of this prior knowledge:

i. Are any of the mechanisms involved in language acquisition particular to the language faculty?
ii. Are any or all of these co-opted from other biological and cognitive systems?
iii. If so, which ones?

Debates *within* the nativist camp, on the other hand, center on questions of the substantive content of UG and the LAD: does UG contain a Sonority Sequencing Principle, and is it violable or inviolable? Does UG provide an invariant initial ranking of markedness constraints (Davidson, Jusczyk, and Smolensky (2004))? Can phonology count (Paster (2012))? And so on. The literature in this area is so enormous that it would be futile to attempt a comprehensive review of it here. For this reason we concentrate on illustrative recent contributions to this debate and organize our discussion around specific themes of current interest rather than taking a historical view of the field. Much of this recent work bears directly on questions (i) and (ii) above, and specifically whether phonology constitutes a counterexample to the idea, associated especially with Chomsky, Hauser and Fitch (2002), that the species-specific component of the human language faculty is rather small, perhaps

consisting only of recursion (see Pinker and Jackendoff (2005) for the argument that it is such a counterexample).

2. Rhetorical typology and practice

Arguments for and against UG and the LAD and their putative components tend to take a limited number of rhetorical forms in phonological contexts. Supporters of UG/LAD typically invoke some combination of cross-linguistic recurrence (e.g. with respect to natural processes in Natural Phonology, or constraints in OT) and gaps (cf. Odden (2005) on retroflexes, or Kiparsky (2006) on final voicing, which we examine in detail in section 4). Proponents of phonological elements in UG have also cited economy considerations; for example Kenstowicz and Kisseberth (1979) and many others have argued that positing syllables as part of the repertoire enables simpler analyses of recurrent cross-linguistic patterns. Learnability considerations have been invoked as well, as with the suggestion that an OT grammar is more easily acquirable because learners need only respond to PLD by ranking a set of constraints provided by UG, rather than constructing rules from scratch as is required in classic RBP (Tesar and Smolensky (1998)). Adherents of OT commonly maintain that Emergence of the Unmarked effects provide evidence for specific universal constraints not motivated by the PLD (McCarthy and Prince (1994)), a variant of the poverty of the stimulus argument. Along similar lines, phonologists working within an RBP framework have suggested that the spontaneous appearance of rules or constraints in first and second language acquisition when unmotivated by evidence from the first or second language may provide evidence for phonological elements of UG such as the Derived Environment Constraint (cf. Eckman, El Reyes, and Iverson (2001) on the acquisition of English by Korean and Spanish speakers) and Identity Avoidance (cf. Vaux and Nevins (2007) on nanovariation in English schm-reduplication).

On the opposite side of the fence, scholars such as Hale and Reiss (2008) have argued that we should not be too quick to attribute cross-linguistic synchronic and diachronic patterns and restrictions to the invisible hand of UG; these may be equally well attributed to independently-motivated properties of the human perception and production systems whose activities extend beyond language per se. Others have added that learning biases implicated in phonological patterns may surface in other animals as well (e.g. Kuhl and Miller (1975) on voice onset time in chinchillas), undermining UG for those who limit its contents to elements specific to humans. It is also possible to mount a plausible argument that at least some typological patterns claimed in the literature are illusory, and that the actual facts of language are more consistent with a model in which for example features are constructed upon exposure to PLD rather than pre-determined by UG (Mielke 2004). Finally, Flemming (2005) and Samuels (2011) have pointed out that UG postulates such as syllables do not always allow for more economical formulations of phonological generalizations, and it is not clear that economy is a valid criterion for deciding membership in UG in the first place.

In order to clarify how exactly the types of argumentation just reviewed are put into practice in UG debates, let us consider in more detail a typical nativist argument that employs the recurrence and gaps arguments: one of the most significant discoveries of 20th century phonology is that the seemingly endless variation in sound patterns in the world's languages is not entirely random. Rather, some phonological patterns and processes are common in many unrelated language families, whereas other imaginable ones are rare or not attested at all. Processes that simultaneously target aspiration and voicing are commonplace, for instance—as in Thai final

laryngeal neutralization (Clements 1985:235)—whereas processes simultaneously targeting aspiration and rounding are not. The same asymmetry holds for repairs of ill-formed phonological configurations; it is often observed for example that aspiration and voicing specifications in syllable coda are typically repaired via devoicing and deaspiration, but rarely[3] via epenthesis (Lombardi (2001); part of the larger Too Many Solutions Problem (Kager (1999), Steriade (2008))).

Such striking facts demand an explanation if valid, and many linguists, especially in the generative tradition, have argued that UG should bear the burden of this explanation. This viewpoint cuts across even deep theoretical divides in the generative literature, being manifest in Rule-Based Phonology from SPE onwards (see especially Chomsky and Halle (1968:ch. 9); Kenstowicz and Kisseberth (1979:251)), and a key tenet of Natural Phonology (Stampe (1979); Donegan and Stampe (1979)) and Classic OT (Prince and Smolensky (1993/2004); McCarthy (2002; 2008)). The following statement from Kenstowicz's standard textbook on Rule-Based Phonology encapsulates this position: "There are many recurrent aspects of phonological structure of a highly specific and rich character whose acquisition cannot be explained on the basis of analogy or stimulus generalization in any useful sense of these terms. These properties are also most naturally explained as reflections of UG" (1994:2).

An alternative view argues that such recurrent aspects of phonological structures can be explained by the interplay of phonetic constraints on articulation and perception and historical change (perhaps in conjunction with general cognitive constraints on learnability). Common phonological patterns and processes reflect common sound changes, and the frequency of a given sound change will depend on articulatory and perceptual factors. This viewpoint is particularly prominent in the work of John Ohala (e.g. Ohala (1971, 1972, 1975, 1981, 2005); Ohala and Lorentz (1977); Chang, Plauché, and Ohala (2001)). It forms the basis of Evolutionary Phonology as pursued by Juliette Blevins in recent work, and has also been adopted by some generativists (Hale and Reiss (2000), Vaux and Samuels (2004), Blevins (2004), Pycha et al. (2003), etc.). As emphasized by Vaux (2003) and Samuels (2011), such a view takes much of the explanatory burden of phonology away from UG and passes it on to phonetics, the language acquisition device, and diachronic linguistics. Samuels (2011) and Samuels, Hauser and Boeckx (this volume), taking this view to its logical end point, argue that this may even allow the formal theory of phonology to be reduced to a small number of primitive operations and categories with counterparts in the cognitive capacities of other species.

The rest of this chapter examines the state of the UG debate with respect to two specific areas in phonological theory where the arguments made are representative of what we find in phonological UG discussions as a whole. Section 3 considers the question of whether UG should account for typological generalizations—that is, whether there should be a direct match between the

[3] Lombardi 2001 and many subsequent publications in Optimality Theory assert that deletion and epenthesis are never recruited to repair violations of constraints on laryngeal specifications in syllable codas, but deletion is attested in Chinese L2 English (Anderson (1983), Xu (2004)) and epenthesis in the L2 English of speakers of Brazilian Portuguese (Major (1987)), Chinese (Eckman (1981), Xu (2004)), Korean (Kang (2003), Iverson and Lee (2006)), and Vietnamese (Nguyen and Ingram (2004)), as well as in first language acquisition (Major (2001)). See Flynn (2007) for further examples and discussion.

phenomena that the theory predicts to exist and those that are attested in the languages of the world. Section 4 reviews the arguments for and against incorporating a notion of segmental and processual markedness into UG, as argued for explicitly in recent Optimality-Theoretic work by de Lacy ((2002, 2006a, 2006b); de Lacy and Kingston (2013)) and Kiparsky (2006; 2008). The evidence and arguments reviewed lead us to conclude that there must be a phonological component of the LAD and perhaps UG as well (in the senses defined in section 1), but that the evidence claimed for specific components of UG in particular is equivocal at best.

3. Constraining the typological space with UG

The desire that the theory of Universal Grammar should predict typological generalizations is present in SPE (see Chomsky & Halle (1968:4-5)) and appears to have been a driving force in the move towards autosegmental phonology (begun in Goldsmith (1976)) in the 1970s and '80s (witness McCarthy's (1988:84) hope that "if the representations are right, then the rules will follow"). More recently, one of the most salient properties of Optimality Theory (Prince and Smolensky (1993/2004)) is that typological predictions emerge automatically from the logically possible permutations of rankable constraints. Since the number of possible rankings of a given OT constraint set is the number of constraints factorial, this property of the model is known as Factorial Typology. Factorial Typology is frequently cited by proponents of OT as a major argument in its favour, as when Féry and Fanselow (2002) state that OT "offers a restrictive theory of linguistic variation: differences between languages can arise only as different rankings of universal principles in different languages" (cf. also Pater (1996) for similar claims).

As an illustration of the workings of Factorial Typology and the use of UG machinery to account for typological generalizations, we take Pater's (1996) work on clusters of a nasal followed by a voiceless stop (which we shall notate as NT clusters henceforth).[4] Pater notes that such sequences are disallowed in many languages, and where they would be created by affixation, phonological processes often eliminate the sequence. The processes that destroy such clusters vary in striking ways from language to language. For example, in many Austronesian languages, the voiceless stop disappears and the preceding nasal appears to assimilate to the place of the deleted stop (this phenomenon is termed "Nasal Substitution". To simplify the exposition, we omit reference to Pater's discussion of root-faithfulness, by which in some languages NT clusters are retained in root morphemes but are eliminated at boundaries between affixes).

(1) Indonesian (Pater (1996:2))
/məN+pilih/ → [məmilih] 'to choose, to vote'
/məN+tulis/ → [mənulis] 'to write'
/məN+kasih/ → [məŋasih] 'to give'

In Mandar, on the other hand, nasals lose their nasality when they precede a voiceless stop:

[4] We employ this case because it is perhaps the one most often invoked by proponents of Optimality Theory, but note that extensive and serious empirical and conceptual problems with Pater's line of reasoning have been pointed out by Blust (2004).

(2) Mandar (Pater 1996:16)
/maN+dundu/ → [mandundu] 'to drink'
/maN+tunu/ → [mattunu] 'to burn'

Still another process is exemplified by Puyo Pungo Quechua, in which NT clusters created by affixation are eliminated via voicing of the stop:

(3) Puyo Pungo Quechua (Pater (1996:21); glosses corrected by NM)

sinik-pa 'porcupine's'	kam-ba 'yours'
saĉa-pi 'in the jungle'	hatum-bi 'in the big one'
wasi-ta 'the house-ACC'	wakin-da 'the others-ACC'

Pater reports that Kelantan Malay exhibits a fourth option, deletion of the nasal and retention of the voiceless stop.

Pater uses the typological fact that NT clusters are often subject to phonological processes, alongside phonetic and acquisitional data concerning these clusters, as evidence that they are marked. In OT terms, this means that there is a constraint in UG which assigns violation marks to NT clusters. Pater calls this constraint *NT:

(4) *NT (Pater (1996:5))
No nasal/voiceless obstruent sequences.

This constraint conflicts with several different faithfulness constraints, each of which favours in different ways the preservation of underlying NT in the surface form. Pater analyses Nasal Substitution as the merger of the stop and nasal, with the resulting segment retaining the place features of the stop and the [nasal] feature of the underlying nasal. Such merger violates a constraint which favours the maintenance of underlying ordering relationships amongst segments. This constraint is known as LINEARITY:

(5) LINEARITY (Pater (1996:9))
S_1 reflects the precedence structure of S_2 and vice versa.

Denasalization, as seen in Mandar, violates a constraint which favours faithfulness to underlying nasality:

(6) IDENT IO [NASAL] (adapted from Pater (1996:17))
Any output correspondent of an input segment specified as [nasal] must be [nasal].

A faithfulness constraint which protects underlying voicing specifications in obstruents militates against postnasal voicing, which is seen in Puyo Pungo Quechua:

(7) IDENT[OBSVCE] (Pater (1996:22))
Correspondent obstruents are identical in their specification for [voice].

Deletion of the nasal, seen in Kelantan Malay, is punished by the MAX constraint in (8).

(8) MAX (Adapted from McCarthy (2008:196))
Assign a violation mark for every segment in the input which has no correspondent in the output.

Pater shows that different permutations of these constraints can model the different processes which affect NT clusters. For example, if *NT dominates LINEARITY, and all other faithfulness constraints also dominate LINEARITY, then Nasal Substitution will result, as shown in the tableau in (9) (numerical indices indicate correspondence relationships amongst input and output segments; the winning candidate is indicated by an arrow):

(9)

/məN$_1$-p$_2$ilih/	IDIO[NAS]	MAX	ID[OBSVCE]	*NT	LIN
→mem$_{1,2}$ilih					*
mep$_1$p$_2$ilih	*!				
mem$_1$b$_2$ilip			*!		
mem$_1$p$_2$ilih				*!	
mep$_2$ilih		*!			

Denasalization will result if IDENT-IO[NASAL] is dominated by all of the other constraints. Nasal deletion results from ranking MAX the lowest, and postnasal voicing by ranking IDENT[OBSVCE] the lowest. Finally, a language in which NT clusters are freely permitted (such as English) is derived by ranking *NT below all of the faithfulness constraints.

This is a classic illustration of factorial typology at work: the attested range of variation is derived by different rankings of universal constraints. However, as Pater points out, a problem arises when one considers the possibility that epenthesis could also eliminate a NT cluster. This would occur if a constraint against epenthesis, known as DEP, were ranked below all of the constraints we have considered thus far:

(10) DEP (Adapted from McCarthy (2008:197))
Assign a violation mark for every segment in the output which has no correspondent in the input.

The result would be as in (11):

(11)

/məN$_1$-p$_2$ilih/	ID-IO[nas]	MAX	ID[OBSVCE]	*NT	LIN	DEP
→məŋ$_1$əp$_2$ilih						*
məm$_{1,2}$ilih					!*	
məp$_1$p$_2$ilih	*!					
məm$_1$b$_2$ilip			*!			
məm$_1$p$_2$ilih				*!		
məp$_2$ilih		*!				

Since DEP is clearly a necessary constraint, Factorial Typology predicts the possibility of the constraint ranking in (11), and hence also the possibility of a language in which epenthesis breaks up NT clusters. The problem is that no such language appears to exist—the Factorial Typology thus predicts the existence of a superset of the actually

attested phenomena. Critics of Optimality Theory (e.g. Hale and Reiss (2000, 2008); Vaux (2003/2008)) take this as evidence for the futility of trying to derive typological predictions from the formal grammar, arguing that what is learnable (sanctioned by UG) is almost certainly a superset of what is attestable, with the latter being explained by independent phonetic, acquisitional, and historical considerations, converging in this respect with work on Evolutionary Phonology (Blevins (2004)), to be discussed below.

Interestingly, some research within the OT tradition has begun to converge on this conclusion as well. Myers (2002) shows that gaps in factorial typologies are "pervasive" (p. 1), arguing that "while there is often more than one attested way of avoiding a marked configuration, it is hardly ever the case that every way of avoiding the configuration is attested […] it is the norm for some of the possible rankings to be unattested". As well as the absence of epenthesis as a way of avoiding NT clusters, Myers points out that metathesis and lenition of the obstruent are also unattested (pp. 6-8). A similar problem besets Myers' earlier (1997) work on the ways in which adjacent identical tones are dealt with in the languages of the world (so-called Obligatory Contour Principle (OCP) effects). While various possibilities predicted by factorial typology are attested, such as tone retraction and deletion, there appears to be no language in which a vowel or a new tone is inserted to break up the OCP-violating sequence (see Myers (2002:8-9) for discussion). Asking why it is that the pervasiveness of these problems had not been noticed previously, Myers attributes it to the narrow scope of much work on factorial typology: "most factorial typologies presented in OT work involve a severely limited set of 'relevant' constraints. It is only when we consider all the faithfulness constraints and all the ways that a representation could be changed to avoid markedness violations that it becomes clear that gaps are in fact the norm". Considering various possible OT responses to this problem and rejecting them, Myers embraces the conclusion that such phenomena, while being allowed by UG, are unattested because "the patterns they represent are unlikely to arise diachronically through natural sound changes on the basis of phonetic patterns." This is precisely the view of Evolutionary Phonology, to which we now turn.

The central premise of Evolutionary Phonology, as Blevins (2004:23) puts it, is that "principled diachronic explanations for sound patterns have priority over competing synchronic explanations unless independent evidence demonstrates, beyond reasonable doubt, that a synchronic account is warranted". The reason for this attitude is that, since a principled diachronic explanation is one which reduces the phenomenon to independent phonetic factors outside of the Language Faculty per se, any statement of the phenomenon at the level of UG constitutes an unnecessary duplication. This reasoning has interesting implications for OT in the light of Myers' (2002) paper: if phonetic and diachronic factors are needed to account for gaps in factorial typologies, then could they also be used to capture the effects of markedness and faithfulness constraints, and if so, is there any more motivation for claiming that the latter are part of UG? We address this question with respect to markedness in Section 4. For now, we illustrate the Evolutionary Phonology approach by discussing how it accounts for the attested and unattested processes which affect underlying NT clusters of the sort discussed by Pater (1996).

In Evolutionary Phonology, all synchronic phonological alternations are the results of sound change. Blevins (2004) follows Ohala (1971, 1972, 1975, 1981, 2005) in identifying the listener as the most important agent in sound change. In other words, frequent sound changes reflect frequent mishearings. Because the NT

configuration is commonly altered by phonological processes, Evolutionary Phonology predicts that phonetic experimentation should show that NT is easily misheard as something else. Moreover, the attested processes which apply to eliminate NT—voice assimilation, Nasal Substitution, nasal deletion and denasalization—should be the results of reasonably common misperceptions by experimental participants in the laboratory. The imaginable ways of eliminating NT sequences which are nonetheless unattested, such as epenthesis and metathesis, should turn out not to correspond to misperceptions of NT sequences uncovered by experimentation.

Myers (2002:13-30) provides articulatory evidence that all of these predictions are plausible, although unfortunately he cites no data from perceptual experiments, which would be the most direct form of evidence for the Evolutionary Phonology position. In particular, he shows through an original articulatory experiment that stops undergo more coarticulatory voicing after a nasal than after a vowel. This adds plausibility to the notion that voiced and voiceless stops are harder to distinguish in postnasal contexts, which could then be the source of a mishearing leading to a post-nasal voicing alternation being introduced into a language. He also shows that in some languages (as documented for Pokomo by Huffman and Hinnebusch (1998)) coarticulation in NT clusters leads to devoicing of the nasal rather than voicing of the stop. Such nasal devoicing obscures the formants characteristic of nasals, potentially leading to the nasal not being perceived at all. In such a case the language learner might either posit that nasal deletion has occurred, or interpret the silence of the nasal as part of the closure for a stop cluster, leading to the denasalization process attested in languages such as Mandar.

The fact that epenthesis and metathesis are not found as repairs for NT clusters can also be explained. Epenthesis usually arises historically either (a) when overlap of phonetic gestures leads to the perception of an inserted segment (Ohala (1983), Browman and Goldstein (1990)) or (b) when the release of a consonant is reinterpreted as a vowel. However, neither of these two scenarios is likely to occur in a nasal-obstruent cluster in a way that is dependent on voicing. If the nasal happened to be released before an obstruent, then this could give rise to epenthesis, but this would affect both NT clusters and clusters of a nasal followed by a voiced obstruent. Meanwhile, a nasal followed by a fricative could give rise to an intrusive stop (as in the common pronunciation of English *prince* as [p^{h}rɪn^{t}s], Clements (1987)), but this produces an affricate, which is still an instance of a nasal-obstruent cluster. Hence, no common phonetic pattern gives rise to epenthesis separating an NT cluster (Myers (2002:25-26)). As for metathesis, Blevins and Garrett (1998) have argued that it arises when the perceptual cues for a particular consonant are sufficiently extended to allow for interpretation as originating in a new position in the string of sounds. In the case of a vowel followed by a glottal stop, the glottalization can extend to the beginning of the vowel, potentially leading to the sequence being interpreted as a glottal stop followed by a vowel, giving rise to metathesis Vʔ > ʔV. Since the cues for obstruency and nasality would not extend in this way, it is extremely unlikely that a process metathesising NT clusters could evolve (Myers (2002:26-7)).

It seems then that Evolutionary Phonology is well equipped to capture the typological data on processes affecting NT clusters without the need for direct reference to innate postulates (even, in the case of factorial typology, improving upon a proposal that does rely on innate postulates). However, a fierce debate has recently erupted concerning whether it is possible to do away with a notion of markedness in Universal Grammar, as we shall now see.

4. Markedness

The notion of markedness in phonology originated with the work of the Prague Circle, most notably in the works of Jakobson and Trubetzkoy, and was taken up in earnest in the central document of early generative phonology (Chomsky and Halle (1968:400-435)). Although the precise definition of markedness varies across different theories, the fundamental content of the notion is always that certain segments, feature values, prosodic structures etc. are in some sense dispreferred—more marked—compared to others. In SPE and its derivational successors, markedness was encoded by a set of markedness conventions which were held to be part of UG. These conventions stated what the unmarked value of a feature was in the context of another group of features. For instance, Convention XXI of SPE (p. 406) tells us that the unmarked value for the feature [±voice] is [-voice] in the context of the feature [-sonorant]; in other words, the unmarked case for obstruents is to be voiceless, rather than voiced. Feature values which were unmarked could be left out of lexical representations and filled in by the markedness conventions. Hence, less marked feature values led to grammars which were more highly valued with respect to economy of lexical storage, and so should be preferred by learners, all else held equal. The markedness conventions were therefore an important plank in the generative approach to capturing typological facts (cf. Section 1). Later, Classic Optimality Theory (Prince and Smolensky (1993/2004)) gave a different formal definition to markedness: a given feature value, prosodic structure or segment is marked if and only if there exists a markedness constraint in UG which assigns violation marks to it. While differing from the classical generative view in allowing markedness constraints to be ranked with respect to each other (and with respect to faithfulness constraints), this view is still one in which the existence and nature of markedness are attributed to Universal Grammar.

An alternative to this is suggested by Myers' (2002:21) succinct statement (although he himself appears to prefer a theory in which markedness constraints are part of UG): "Marked phonological structures are those that tend to be misheard as something else." The view that synchronic markedness can in fact be reduced to this in conjunction with diachronic change has been pursued in much work by Ohala, Blevins and certain generativists (viz. Ohala (1971, 1972, 1975, 1981, 2005); Ohala and Lorentz (1977); Chang, Plauché, and Ohala (2001); Hale and Reiss (2000); Vaux and Samuels (2004); Blevins (2004)). We have already seen from the case study of NT clusters in Section 3.1 that this approach is successful in accounting for certain instances of markedness as it effects typology (Blevins (2004) catalogues many other examples). However, there has been some debate as to whether this approach is sufficient to capture all aspects of markedness. Many objections have been raised in the work of de Lacy and of Kiparsky (de Lacy 2002, 2006a, 2006b; de Lacy and Kingston 2013; Kiparsky 2006, 2008), to which we now turn.

4.1. Velar epenthesis

Some of de Lacy's and Kiparsky's arguments for the need to posit markedness constraints in UG depend on examples of marked configurations which are plausible results of sound change but which are claimed to be unattested and impossible. For example, de Lacy and Kingston (2013) assert that phonological velars are never epenthetic and are never the outcomes of place neutralization, in spite of static distributional patterns in some languages suggesting that (for instance) velar neutralization has occurred as a sound change. De Lacy and Kingston (2006:293-294)

illustrate this last pattern with Peruvian Spanish forms like those in (12) and (13) below; several similar cases are discussed in Howe (2004) and de Lacy and Kingston (2013).

(12) Peruvian Spanish static pattern: stop codas in loan words are velar
pepsi [peksi]
Hitler [xikler]

(13) Peruvian Spanish (some speakers)
apto 'apt' [akto] (homophonous with acto 'act')
abstracto 'abstract' [akstrakto]
opcional 'optional' [oksjonal]

Howe (2004:17) mentions that Cuban Spanish possesses not only a similar static pattern of coda stops being velar but also related alternations including the prefix *sub-* having velar realizations like [suk]- before a consonant vs. [suβ]- before a vowel. In order for this not to constitute an example of synchronic velar neutralization, it would be necessary to insist that the alternation is not phonological. Yet since velar preconsonantal codas are predictable in this dialect, such an insistence seems motivated by nothing other than the desire to preserve the putative universal that velar neutralization does not occur as a synchronic phonological process.

Rice (1999) and Howe (2004) also mention Uradhi dialects reported with a word-final epenthetic segment realized as [ŋ] if the next consonant leftward is nasal, otherwise variably as [k] or [ŋ]. While acknowledging uncertainty about this example, De Lacy and Kingston (2013:309) suggest that the [k] allophone shows long-distance consonant harmony. Synchronically, however, it seems possible that the invariably nasal epenthetic allophone is the one harmonizing, and such harmony would only be long-distance if the intervening vowel were phonetically oral. De Lacy and Kingston further suggest that the supposedly long-distance-harmonizing [k] might be a phonologically predictable, semantically meaningless reduplicant morpheme. They mention Gafos (1998) treating some long-distance consonant harmonies as instances of reduplication, along with Rose and Walker (2004) modelling long-distance consonant harmonies with correspondence constraints like those that have been applied to reduplication in OT, but of course it does not follow that all cases of such harmony are reduplicative (cf. Hansson (2001)). Further, the agreement-based analysis of consonant harmony encounters a number of serious problems that do not arise for an alternative using relativized locality (Nevins and Vaux (2004)).

De Lacy and Kingston (2013) effectively dismiss the remaining kinds of evidence in Howe (2004) as amounting to nothing more than sound changes possibly involving epenthesis of velars or neutralization to velars but not producing synchronic phonological patterns of velar epenthesis or neutralization. However, additional arguments that [k]-epenthesis occurs as a synchronic pattern are marshalled by Vaux and Samuels (2012) (cf. Blevins (2006b:253) citing Vaux (2003)) in support of a broader claim that any segment can become epenthetic as the result of reanalysis of a historically earlier process deleting that segment in some environments. One of the better-known examples of this is the [r]-insertion found in many non-rhotic varieties of English, a hiatus prevention strategy derived by reanalysis of earlier coda [r]-deletion.

4.2. Final voicing

In a similar controversy, Kiparsky has drawn attention to the dearth of examples of obstruents neutralizing to voiced segments in syllable- or word-final position, a process we hereafter term "final voicing". Plausible phonetic motivations for word-final devoicing are well-known (see Blevins (2004:103-106) for a summary). However, Kiparsky (2008:45) asserts that final voicing does not occur. He argues that it could easily arise through several ordered sets of possible sound changes; thus, its non-occurrence must be due to a synchronic constraint somehow ruling out such diachronic scenarios, and Evolutionary Phonology is inadequate because, he argues, it lacks the necessary type of constraint. Kiparsky (2008:47) suggests two diachronic scenarios that would produce final voicing, shown in (14a-b). Kiparsky (2006:223) adds the scenario shown in (14c) (along with two others treated below in (23)).

(14) Final voicing scenarios

a. Scenario 1: chain shift resulting in markedness reversal
Stage 1: tatta tata tat (*tatt) (gemination contrast)
Stage 2: tata tada tad (*tat) (lenition)
• Result at stage 2: new voicing contrast, word-final phonological voicing.

b. Scenario 2: lenition plus apocope
Stage 1: takta tada (*tata, *data, *tat, *dat) (allophonic V___V voicing, no final -C)
Stage 2: takta tad (*tat, *dat, *dad, *dat) (apocope, unless final *-CC would result)
• Result at stage 2: allophonic voicing of word-final stops.

c. Scenario 3: lenition plus deletion
Stage 1: tat tad dat dad (voicing contrast)
Stage 2: tad tað dad dað (coda lenition)
Stage 3: tad ta dad da (loss of weak fricatives)
• Result at stage 3: only voiced stops occur in codas.

Even if these scenarios did lead to a system with final voicing, an important caveat must be sounded with respect to such diachronic arguments in general. In order for them to be convincing cases of overgeneration by the Evolutionary Phonology model, it must first be shown not only that each individual stage has a non-negligible probability of occurrence, but that the probability of all stages occurring ($=P(S_1)\times P(S_2)\times P(S_3)$, assuming the changes are independent) is not vanishingly small. Otherwise, the apparent absence of such systems in attested natural languages could reduce to the unlikeliness of the whole scenario taken together, and would thus not constitute a challenge to Evolutionary Phonology. Unfortunately, it is not clear at this point how to establish the likelihood of these scenarios without making arbitrary assumptions about the probability of occurrence of each individual stage.

Even if a way around this obstacle can be found, it seems to us that the scenarios in (14) produce postvocalic voicing rather than final voicing (see also Blevins (2006a:145), who makes this observation for the first two scenarios). To further confirm this point, we examine one language not mentioned in this debate— Middle Persian— which appears to offer an approximate equivalent to a combination of Kiparsky's second and third final voicing scenarios (see 14a-b). In what follows, we use the Middle Persian facts as a means of unpacking the content and predictions of Kiparsky's model.

Proto-Iranian possessed a two-way laryngeal opposition in obstruents word-initially (15a); this contrast was preserved in Middle Persian (here represented by transcriptions of the written form known as Pahlavi) (15b).

(15) selected[5] word-initial laryngeal developments in Iranian

	a. P-Ir	**b. Pahlavi[6]**	**c. Avestan[7]**	**gloss**
*b-	*brātā-	brād	brātā	brother
*d-	*dāru-	dār	dāru-	wood
*g-	*gāu-	gāw	gau-	cow
*p-	*pitar-	pidar	pitar-	father
*t-	*tanū-	tan	tanū-	body, self
*k-	*kapauta-(ka-)	kabōd	OP kapautaka	grey-blue

One can also see in (15b) that by the time of Pahlavi, Proto-Iranian final syllable rimes were lost in polysyllabic words[8]; compare for example P-Ir. ***dāru-*** 'tree' with its Pahlavi descendant *dār*. This process of rime loss can be compared to Stage 2 of Kiparsky's Scenario 2 in (14).

Proto-Iranian also contrasted voicing in stops in post-sonorant position (16a), but Pahlavi lost this contrast as a result of spirantization of original voiced stops in later Old Iranian (represented by the Young Avestan examples in (16ic); compare Stage 2 of Kiparsky's Scenario 3) followed by Pahlavi voicing of original voiceless stops in this environment (see (16iib)), probably by the fourth century AD and certainly by the ninth century (Mackenzie (1967, 1971)).

(16) selected medial laryngeal developments in Iranian

	a. P-Ir	**b. Pahlavi**	**c. Avestan**	**gloss**
i.	*b *ābar-	āwar-	āβar-	bring
	*d *madaka-	mayg	maðaxa-	locust
	*g *yuga-	ǰuɣ	yaoga-	yoke
ii.	*p *āpa-	āb	āpa-	water
	*t *āzāta-	āzād	āzāta-	free, noble
	*k *parikā-	parīg	pairikā-	witch, demoness

Iranists such as Pisowicz (1984) and Sims-Williams (1996) have suggested that the spirantization and voicing developments are connected, with the loss of voiced stops in this position allowing for the voiceless series to move into the vacated slot in a

---

[5] We omit the palatal affricates here, as their situation is more complicated (q.v. Pisowicz (1984:22)).

[6] All Pahlavi forms are from Mackenzie (1971).

[7] We provide cognates from Old Iranian languages (Avestan or Old Persian (OP)) to give the reader a better sense of why Iranists assume the particular Proto-Iranian forms presented here.

[8] Iranists tend to label this process as final syllable loss (e.g. Schmitt (1989:60), Weber (2007:942), Korn (2009:207)), missing the facts that the onset of the final syllable normally does not delete and that monosyllables do not undergo the same reduction.

context where voicing is phonetically favored (i.e. after vowels, liquids, nasals, and perhaps voiced fricatives).

Iranists typically formulate the voicing process reflected in (16iib) as happening intervocalically (Korn (2003:55), Paul (2008)), postvocalically (Back (1978) and Korn (2009:208)), or after voiced continuants (Weber (1997:613)), rather than after voiced segments, the generalization we employ here. It is difficult to determine whether the process placed any restrictions on what could immediately follow the target consonant, i.e. whether it required a following vowel or could apply in word-final position, because we lack Pahlavi forms derived from Proto-Iranian forms ending in final voiceless consonants. We can refine the statement of what could immediately precede the targeted consonant, though, on the basis of forms of the sort in (17); the intervocalic and postvocalic generalizations mentioned above fail to account for the voicing that occurs after consonantal sonorants[9] (as well as voiced fricatives according to Pisowicz (1984:18), though he provides no examples), and the post-continuant theory makes the wrong predictions for nasal contexts (since nasals are [-continuant] yet trigger voicing).

(17) Pahlavi post-consonantal voicing

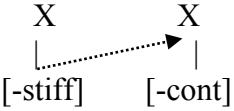
	a. P-Ir	**b. Pahlavi**	**c. Avestan**	**gloss**
*nT	*spanta-	spandarmad	spəṇta- + ārmaiti	Holy Thought
	*antar(-aka)	andar(ag)	aṇtarə	between
	*panča-	panǰ	paṇča	five
*rT	*martya-	mard	martya-	mortal, man
	*krp-	kirb	kəhrp-	body, form
	*wrka-	gurg	vəhrka-	wolf

Stops do not voice after voiceless segments; cf. P-Ir. **hapta-* 'seven' > Pahlavi *haft*; **wispa-* 'all' > Pahlavi *wisp*, **wist* 'twenty' > Pahlavi *wīst*.

In light of these data, we propose to formalize the voicing process with the general assimilation rule in (18).

(18) Pahlavi assimilation of stops[10] to a preceding voiced segment

[-cont] → [-stiff vocal folds] / [-stiff] _



But was it the fact that the voiceless stops followed a voiced segment that led them to undergo voicing, or the fact that they were in final position following the process of final rime loss mentioned earlier? If the latter, we might be approaching something like Kiparsky's final voicing scenario. Most of the Pahlavi forms presented thus far are compatible with both possibilities, but *pidar*, *kabōd*, *spandarmad*, and *andarag*, wherein voicing applies word-internally, are not. Forms like these and the ones in (19)

[9] This process appears to have been blocked by at least some morphological boundaries (cf. Weber (1997:613)); e.g. **ham-prsa-* 'consult', lit. 'ask together' > Avestan *hampərəsa-* 'to deliberate', Pahlavi *hampursag* 'consulting' (not **hambursag*).

[10] Note that by the formulation in (18), the rules applies vacuously to nasals as well.

demonstrate that the voicing process in Pahlavi must have been triggered by a preceding voiced segment, not by the fact of being in word-final position.

(19) relevant forms with intervocalic stops in Pahlavi

	a. P-Ir	**b. Pahlavi**	**c. Avestan**	**gloss**
*p	*xšap-ika-	šabīg[11]	xšap-	darkness, night
*t	*dātār-	dādār	dātāram (acc. sg.)	creator
*k	*wikaya-	gugāy	vīkaya-	witness

The Pahlavi facts present a situation of approximately the sort outlined by Kiparsky (2006) in Scenarios 2 and 3; we schematize the parallelism in (20).

(20) Scenarios 2 and 3 and their Iranian parallel

Scenario 2	**Iranian**	**Scenario 3**
	voicing contrast (15a, 16a)	Stage 1
--	post-vocalic spirantization (16ci)	Stage 2
Stage 2	final Rime deletion	--
Stage 1	post-voiced voicing (16bii)	--

Result: post-voiced (and thus most final) stops are voiced; conditions are met for learners to postulate a post-voiced voicing process whose targets include coda stops.

To paraphrase Kiparsky (2006), what we see in the historical development of Pahlavi is a scenario that could produce the effect of a synchronic rule of post-voiced voicing. But did speakers of the Middle Persian variety represented in writing as Pahlavi actually postulate such a synchronic rule, or did the process die as soon as it had applied? The limited nature of the corpus makes it difficult to address this question, but a number of suggestive facts can be identified.

The fact that Avestan liturgical loans such as *ātaxš* 'fire' (in contrast to the native outcome of the same root, *ādur*) do not undergo the voicing process (cf. Weber (1997:613)) might seem to suggest that rule (18) was not synchronically active in Pahlavi, at least by the time these words were borrowed. However, we know from the work of Pierrehumbert (2006) and others that synchronic processes can remain active without applying to new forms, so the exceptional Avestan loans do not prove that rule (18) was not synchronically active in Pahlavi Middle Persian.

Indeed, the existence in the language of extensive voicing alternations resulting from morpheme concatenation suggests that the voicing rule remained at least partially productive in the synchronic grammar of Pahlavi. Consider for example the suffixes in (21):

(21) voicing alternations in Pahlavi (affixed forms from Weber 2007)

affix		**affixed form**	**gloss**
infinitive –tan < PIr *-tanaiy	a.	xuf-tan	sleep
		raf-tan	go
		kuš-tan	kill
		abrōx-tan	illuminate

[11] Mackenzie (1971) glosses this as 'Mazdean ritual undershirt'; compare the Armenian loan *šapīk* 'shirt'. Originally it presumably meant something like 'thing worn at night'.

		rēx-tan	flow
		bas-tan	bind
	b.	pursī-dan	ask
		dā-dan	give, create
		abzū-dan	increase
		bū-dan	be(come)
		ēstā-dan	stand
		kar-dan	do, make
		xwar-dan	eat
		za-dan	hit
		kan-dan	dig, destroy
past –t < PIr *-ta-	c.	āwiš-t	sealed
		kar-d	made
		dā-d	gave, given
		guf-t	spoke(n)
superlative –tom < PIr *-tama-	d.	wat-tom	worst
	e.	abar-dom	highest
		ab-dom	last, final
		bē-dom	furthermost
		fra-dom	first
		di-dom	second
		ni-dom	least, smallest
agentive –tār < PIr *-tar-	f.	bōx-tār	savior
		dāš-tār	keeper
		guf-tār	speaker
		kas-tār	destroyer
	g.	amenī-dār	unthinking
		bur-dār	bearer, womb
		dā-dār	creator
		dī-dār	sight
		framā-dār	commander
		gā-dār	husband
		handēšī-dār	thoughtful
		kar-dār	worker

The reader might be objecting at this point that one could equally well postulate underlying representations with initial voiced stops for the affixes in (21), which in tandem with a rule spreading the voicing value of a segment to a following stop would yield the desired surface forms and obviate the need for a voicing rule of the sort in (18). A number of considerations suggest that a devoicing analysis of this type is not to be preferred. First, suffixes beginning with original voiced stops (e.g. *-bān* ‘keeper of X’, *-bed* ‘lord’) do not show voice alternations, as we would expect them to if the grammar contained a devoicing rule of the sort described above. Moreover, the superlative form of *wad* ‘bad’, *wattom* ‘worst’ (21d), works nicely if we assume underlying forms /wad/ and /-tom/ and a garden variety voicing assimilation rule ordered before rule (18) that devoices the /d/ before the following /t/, but makes little sense if the underlying form of the suffix is /-dom/—in this case we should expect the superlative to be *[waddom].

A potential problem is posed by a small number of affixes that have non-alternating initial voiceless stops, including the comparative *-tar* < **-tara-*.

Derivations such as *wad* 'bad' → *wattar* 'worse' look at first blush exactly parallel to the above-mentioned *wad* 'bad' → *wattom* 'worst', suggesting underlying forms /wad/ and /-tar/. Sonorant-final roots reveal that the situation is somewhat different, though; contrast for example *abēr-tar* 'much more' with *abar-dom* 'highest', *abzōn-tar* 'more increasingly' with *kan-dan* 'dig'.

Pahlavi thus presents a three-way contrast in voicing behavior, with non-alternating voiced stops (*-bed* 'lord'), non-alternating voiceless stops (comparative *-tar*), and stops that undergo voicing alternations (superlative *-tom* ~ *-dom*). This situation is reminiscent of what we find in Turkish, which possesses a contrast in stem-final position between non-alternating voiced stops (*etüd* 'étude' : *etüdüm* 'my étude'), non-alternating voiceless stops (*at* 'horse' : *at-ım* 'my horse'), and alternating stops (*at* 'name' : *ad-ım* 'my name'). Inkelas and Orgun (1995) and Hale and Reiss (2008) propose to analyse ternary oppositions of this type in terms of equipollent features combined with underspecification; the Turkish non-alternating /d/ for example would be underlyingly specified as [-stiff], the non-alternating /t/ as underlyingly [+stiff], and the alternating *d~t* as underlyingly unspecified for [stiff].

This sort of archiphonemic analysis can be extended to the Pahlavi case, but requires certain modifications (for both Turkish and Pahlavi) in order to account for the behavior of sequences of alternating consonants. By the logic applied in the Turkish case, the [b] of Pahlavi *-bed* should be underlyingly [-stiff], the [t] of *–tar* should be underlyingly [+stiff], and the [d]~[t] of *-dom/-tom* and of *wad-/wat-* should be underlyingly unspecified for [stiff]. The problem here is that a form like /waD-Dom/[12] should then involve two adjacent coronal stops each unspecified for [stiff vocal folds], and it is unclear why they should be subsequently specified as [+stiff] in the surface form [wattom], rather than taking the [-stiff] specification of the preceding vowel, as is the case in /waD/ 'bad' → [wad]. Interestingly, the same problem arises in Turkish, as shown for example by alternating suffixes like ablative /-DAn/ (cf. *at-tan* 'horse-ABL' vs. *adam-dan* 'man-ABL'), which produce voiceless clusters when affixed to stems ending in an alternating stop, e.g. /kitaB/ 'book' → [kitap] 'book-NOM', [kitaba] 'book-DAT', [kitaptan] 'book-ABL'.

We propose (relatively uncontroversially) that underspecified segments which do not receive a value for the feature in question from the application of other phonological rules (as the /D/ in Pahlavi /waD/ would from the preceding /a/ by rule (18)) are assigned the unmarked value for that feature ([+stiff] in the Pahlavi case) by default. But why does /waD-Dom/ receive [+stiff] by dint of this redundancy rule rather than [-stiff] from the preceding /a/ by application of rule (18)? Applying the redundancy rule before rule (18) would yield the desired result when combined with a leftward rule of voicing assimilation in stop clusters of the sort mentioned earlier. Two possibilities in the phonological literature may be applicable here. First, one could propose that rule (18) is preceded by a rule that coalesces strings of adjacent identical segments into linked structures, which could then lead to the cluster resisting application of rule (18) due to the Uniform Applicability Condition that has been claimed to produce geminate inalterability effects in numerous languages (Schein and Steriade (1986)). A second possibility involves the fact that voiced geminates are said to be aerodynamically marked due to the longer closure increasing the probability of voicing failure (Hayes and Steriade (2004:6ff)). In other words, the lack of *[waddom] can be captured with rule ordering, or related to a broader formal

[12] We henceforth use capital letters to denote archiphonemic representations underlyingly underspecified for the feature under discussion.

principle (geminate inalterability), or phonetically grounded (aerodynamics of voicing), illustrating competition between some of the same kinds of explanations (parochial, general-formal, and functional) that are discussed more broadly in the rest of this chapter.

To summarize our discussion of Pahlavi, we have seen that the language appears to have followed the sort of historical trajectory that we find in Kiparsky's second and third scenarios (see 14b,c), and that this trajectory yields not the cross-linguistically rare final voicing that Kiparsky suggests, but rather a common and natural synchronically productive rule of voicing assimilation that spreads [-stiff vocal folds] from a voiced segment to an immediately following stop, both word-internally and word-finally.

The fact that Kiparsky's first two scenarios produce postvocalic voicing rather than final voicing was already noticed by Blevins (2006a:145). The same can be said of Kiparsky's third scenario, where incidentally the final step of fricative loss is superfluous in that it does not help produce a pattern of having only voiced obstruents in codas.

To see why these scenarios either fail or are unlikely to produce synchronic rules of final voicing, let us first consider what would happen to suffixed forms in Scenario 1. Let us suppose that this language contained one or more vowel-initial suffixes (a cross-linguistically common state of affairs in human languages), one of which we will assume for ease of exposition to be /-a/. Kiparsky's proto-form **tat* would then have the suffixed counterpart **tat-a*. This pair should yield {*tad, tada*} at Stage 2, following application of what Kiparsky appears to assume is a post-vocalic lenition rule. Learners exposed to paradigms of this type would most likely postulate underlying forms /tad/ and /a/ for the relevant morphemes, with no rule of final devoicing; this simple system would directly generate the desired surface forms. If on the other hand they were to postulate /tat/ and /-a/ plus a rule of final voicing, the grammar would require an additional rule to generate the voicing in suffixed [tada]. One could imagine an aggressive learner postulating a rule of the latter type to account for the static fact that post-vocalic stops are invariably voiced in this language, but such a rule would obviate the need for a rule of final voicing.

One could hope that learners would induce a rule of final voicing if there were multiple suffixes consisting only of geminates which had reduced to singletons in word-final position in the parent stage, since in the daughter stage these suffixes would show voiceless and voiced singleton stops in word-medial and final position respectively. Yet it seems suspicious to make the final voicing rule rely on such suffixes: there could not be very many of them, and consequently their alternations might be learned without any phonological rule at all. The two more plausible grammars for Scenario 1, then, are one with no voicing rule at all, and one with a general rule of post-vocalic voicing. In either case, it is highly unlikely (and perhaps impossible) that a rational learner would postulate a voicing rule restricted to final position.

Scenarios 2 and 3 force a similar conclusion. When one considers words of more than two syllables, it becomes clear that Kiparsky's scenarios would produce post-vocalic voicing, not final voicing. Consider the sample forms in (22):

(22) Outcomes including longer words in Kiparsky's Scenarios 2 and 3
a. Scenario 2

- Stage 1: allophonic V___V voicing, no final –C
    - takta, tada, tadada

- Stage 2: apocope, unless final *-CC would result
    - takta, tad, tadad

b. Scenario 3

- Stage 1: voicing contrast
    - tat, tad, dat, dad, tatat, tadad…
- Stage 2: coda lenition
    - tad, tað, dad, dað, tadad, taðað…
- Stage 3: loss of weak fricatives
    - tad, ta, dad, da, tadad, taa (?)…

Here the key forms are original *tadada* in (22a) and *tatat* in (22b). The former emerges from Stage 2 as *tadad*, showing that stops are voiced not word-finally, but post-vocalically; the exact same generalization is shown by *tatat* in Scenario 3, which emerges from Stage 3 as *tadad*.

Blevins (2006a:144-153) suggests that six languages developed what amounts to a postvocalic voicing pattern and that Evolutionary Phonology is consequently right on target in postulating that UG lacks any constraints forbidding this pattern. Kiparsky (2006) in turn rejects all these examples. Here we review the ones in which further details, parallels, or corrections seem worth mentioning. We agree with Kiparsky that none of the six languages clearly support Blevins' (2006) claim, but we argue that this fact is not necessarily a problem for Evolutionary Phonology.

A number of ancient languages are thought by some writers to have undergone neutralization of post-sonorant word-final stops to voiced forms, including Old Latin (see Meillet and Vendryes (1968:146), Weiss (2009:155)). Proto-Italic is thought to have changed earlier postvocalic final *-*t* to *-*d*, which survives in a number of early inscriptional forms in various Italic languages (Sihler (1995:228-229)). Kiparsky (2006:230-232) suggests based on wider developments within Italic that this segment may have been a voiced fricative when (if ever) its divergence from /t/ constituted a synchronically active phonological pattern, but this does not necessarily affect its relevance to the question of final obstruent voicing. Certainly relevant though is Blevins' observation that "while there are only a few morphemes for which the sound change is attested, there is no evidence in this case against a general final obstruent voicing process" (2006a:145-146). The basic problem here is that the data are consistent with Old Latin having undergone a process of final voicing, but the amount and type of data are insufficient to determine whether the process was synchronically active, and whether or not it extended beyond the few coronal cases for which it is attested.

In passing we note even more severe qualifications on the possibility of detecting final voicing in Hittite (see Melchert (1994:18)) and in Proto-Indo-European itself (see Ringe (2006:143)). First, any final voicing rule in Proto-Indo-European would probably need to exclude unsuffixed monosyllables; a voicing contrast in monosyllable-final stops is suggested by the relatively secure etyma **k̂erd* 'heart' vs. **yek^{w}r̥t* 'liver', reflected in Classical Armenian as *sirt* and *leard* respectively (thanks to James Clackson for pointing this out). Second, evidence for a synchronically active final voicing pattern in Proto-Indo-European seems limited to a few morphemes such as 3sg *-*t*- and 3pl *-*nt*-, which would have been voiced word-finally but voiceless when followed by the "hic et nunc" –*i* (thus **b^{h}éreti* 'is carrying', **b^{h}éronti* 'are carrying' with primary endings, vs. the non-*hic-et-nunc* equivalents **b^{h}éred*, **b^{h}érond*, which have secondary endings; forms from Ringe (2006:143)). Within this limited set of data, the inference of final voicing itself depends on scant evidence like

that of Italic already mentioned. Finally, in Hittite, the non-place stop contrast is recorded orthographically through somewhat inconsistent gemination, which has been interpreted in a variety of ways besides a voice distinction (q.v. Melchert (1994)).

Welsh is the next candidate for final voicing in Blevins (2006a:146-147). Kiparsky (2006:227-228) counter-claims that the stop contrast in this language is one of aspiration with variable phonetic voicing of the unaspirated series. Even if the phonological contrast involved [stiff vocal folds], though, the Welsh data would resist a final or even post-voiced or post-sonorant voicing analysis. Blevins admits that the language has final voiceless stops but observes that in monosyllables where the rhyme is a vowel followed by a stop, the vowel is predictably short if the stop is one of /p/, /t/, or /k/ and long if it is one of /b/, /d/, or /g/. In fact, vowels in stressed monosyllables are predictably long before coda obstruents other than /p/, /t/ and /k/ (see Wood (1988:231-232)), so any analysis positing final voicing after long vowels would have to limit it to stops. Though this might work, rather than entertaining the possibility, Blevins (2006a:147) claims that the important fact here is counterfactual: if not for loans ending in /p/, /t/, and /k/, Welsh would have had final voicing. Kiparsky (2006:228) understandably responds that real synchronic systems are precisely the topic here, not imaginary alternatives. However, words ending in /p/, /t/, or /k/ are not all loans; after nasals these segments are a regular native reflex except that the coronal stop has disappeared in clitics and variably in polysyllables. Examples include *dant* 'tooth', *pump* 'five', *ieuanc* 'young' in both Early Welsh (Strachan, Meyer, and Lewis (1937:8)) and Modern Welsh (Evans (1852) and University of Wales (2012)). Further, British word-final *–lt(V(C))* yielded Welsh *–llt*, while otherwise an earlier post-liquid voiceless stop onset of a final syllable generally yielded a post-liquid word-final voiceless fricative in Welsh (see Schrijver (1995:349)). This shows that Welsh underwent no sound change of final voicing per se, but at most only postvocalic or intervocalic voicing, assuming that the stops contrasted in voice rather than aspiration when this took place. We are unaware of evidence that this historical voicing process led to any synchronic phonological pattern of active voicing.

The next candidate for final voicing is Somali (Blevins (2006a:147-148)), which has two contrastive stop series, one aspirated and the other unaspirated. Unaspirated stops at the bilabial, coronal, and velar places of articulation are medially voiced between voiced segments; except for the lone (coronal) implosive, they are also spirantized between vowels unless the second vowel is a phonetic schwa offglide as described further below (Edmondson, Esling, and Harris (2004), cf. Gabbard (2010:20-22)). Word-initially, unaspirated stops other than the glottal and epiglottal-uvular are described by Edmondson, Esling, and Harris (2004) as partly voiced, with at least the bilabial being entirely voiceless in the speech of one informant. Gabbard (2010:7-11) shows 86-115 ms of voicing for the bilabial, non-implosive coronal, and velar unaspirated stops in apparently utterance-initial position (i.e. preceded by a flatline waveform), so perhaps the degree of word-initial voicing varies considerably by speaker.

Gabbard (2010:7,10) generalizes that non-uvular voiceless stops are aspirated, but without providing any argumentation for voice in coda stops. Edmondson, Esling, and Harris (2004:2-5) go into more detail. At the end of a word (perhaps when not followed by a vowel-initial word, though this is not stated), unaspirated stops other than the glottal stop are followed by a schwa offglide "in careful, overly correct speech", with non-uvular ones being voiced. In the same environment in "conversational style", stops apart from the implosive are voiceless glottalized; the

examples are all unaspirated, and it is stated that aspirated stops are not found in codas. Coda unaspirated stops as the first member of a word-medial geminate are also identified as voiceless glottalized. This last point disagrees with Gabbard (2010:14, 28-29), who transcribes the geminates phonetically as voiced singleton stops but provides no experimental evidence on either closure duration or uniformity of closure voicing.

On the whole, regardless of how the laryngeal contrast in Somali stops is phonologically specified, it appears that only unaspirated stops occur in codas, that they are voiceless glottalized in ordinary speech (at least of Edmondson et al.'s consultants), and that in especially careful speech, some underlyingly final stops are voiced with a schwa offglide. The fact that underlyingly final epiglottal-uvular stops are followed by this schwa without being voiced makes it harder to argue that the schwa is merely a result of sustaining underlying voice through stop release. Conceivably, the schwa is an artefact of stop release itself in careful speech, and the more anterior stops have become voiced before it as a result of being phonetically intervocalic and prone to greater degrees of voice leak from a preceding vowel. (On the aerodynamic correlation of voice with anteriority see Hayes (1997), Helgason and Ringen (2008), Ohala (1997).) This would entail that spirantization is more restricted than voicing since it would apply only between underlying vowels.

Leaving the treatment of Tundra Nenets to Kiparsky (2006), we turn now to Blevins' (2006a) last example of putative final voicing. Lezgian (Yu 2004) has an alternation between plain voiceless and voiced stops where the voiceless alternant occurs in onsets and the voiced one in codas at the end of a large set of monosyllabic stems (these codas are word-final unless followed by suffixes or by the second element of a compound). The alternating stops contrast with non-alternating voiced stops, voiceless aspirates, and ejectives, all of which are found in both onsets and codas including monosyllabic word-final position. The question is how to represent the alternating pair phonologically. If the onset alternant is taken as basic (as in Blevins (2006a:150-152) and Yu (2004)), then Lezgian has a pattern in which otherwise plain voiceless stops are voiced in codas.

Kiparsky (2006) instead takes the coda allophone as basic and underlyingly geminate, treating the alternation as a case of onset devoicing and degemination. Yet while the coda alternant does appear to be the historically more conservative one, it is not clear whether Lezgian learners would consider it either underlying or geminate. As seen in Yu (2004), its closure duration is about a quarter longer than the duration of its voiceless intervocalic onset alternant, about a third longer than onset non-alternating voiced stops, and about a fifth longer than coda non-alternating voiced stops. Would these length differences provide a sufficient basis for treating the coda alternant as geminate while treating all the other sounds just mentioned as singletons? Kiparsky notes that onset devoicing occurs in other languages but does not provide examples where voiced or any other kind of geminate stops surface only when not followed by a vowel. In fact, Yu's (2004) historical analysis is that the coda alternants are and were singletons and that they geminated in onsets (which for independent reasons were generally pretonic), subsequently devoicing and then degeminating in Lezgian.

The closure and voicing duration differences between alternating and non-alternating coda voiced stops – 25 and 34 ms average in the tokens measured – shows that they do not completely neutralize (Yu (2004:81-83)). For a critique of research on "partial" neutralization see Yu (2011), who favors treating it as a decrease in the set of cues available to express a contrast in particular environments rather than necessarily

involving actual categorical neutralization in the phonological domain (while also discussing factors that have motivated analyses of the latter type). If Lezgian does not involve categorical neutralization of alternating and non-alternating coda voiced stops, then it is not a counter-example to Kiparsky's claim that such neutralizations are non-existent.

It would still involve final voicing of an underlyingly voiceless or laryngeally unspecified stop series if the onset alternants are taken as basic, but again it is unclear which alternant learners select as basic. Yu (2004:76-78,87-88) notes that Lezgian has additional lexically restricted alternations between prevocalic ejectives and word-final voiced stops or aspirates. The word-final voiced alternants of prevocalic ejectives are virtually identical in closure and voicing duration to those of prevocalic plain voiceless stops (the final alternants of the prevocalic ejectives average 7 ms longer in closure duration and 10 ms shorter in voicing duration than those of the plain voiceless stops in the tokens measured; Yu (2004:81)). At the same time, the restriction of both voicing alternations (with plain voiceless and ejective onsets) to particular sets of monosyllables within Yu's data could mean that neither alternation is synchronically productive (Yu (2004:93)); monosyllables with non-alternating voiced stop codas include both obvious loans and less easily explained forms, and Yu does not discuss other factors which might indicate which patterns are currently productive (see Yu (2004:89-92)). Thus we remain open to the possibility that the putative phonological neutralization to voice in Lezgian codas is neither productive, nor a neutralization, nor a process with a voiced outcome.

To summarize the discussion so far, Evolutionary Phonology does not predict that final voicing may be prohibited—merely that its likelihood of occurrence depends on sound change; to date, no single sound change is known to produce final voicing, and hypothetical sequences of sound change like the ones in (14) generate not final but intervocalic or postvocalic voicing patterns (Blevins (2004:108-110; 2006a:16,145)). We have seen that synchronic postvocalic or perhaps coda voicing may exist in Somali careful speech and in Lezgian, though the details also seem open to other interpretations.

Besides the three diachronic scenarios in (14) which lead at best to postvocalic or post-voiced voicing patterns rather than final voicing proper, Kiparsky (2006: 223-224) offers two other diachronic scenarios capable of producing synchronic final voicing patterns – but arguably at some cost to plausibility. These are shown in (23).

(23) Kiparsky's fourth and fifth final voicing scenarios
a. Scenario 4: assimilation plus deletion
Stage 1: tata tanta (no voicing contrast, only nasal codas)
Stage 2: tata tanda (allophonic voicing after nasals)
Stage 3: tata tand (apocope after heavy syllables)
Stage 4: tata tad (loss of nasals before stops)

Stage 2 is like Japanese. At Stage 3, final vowels are lost after heavy syllables, as in Old English. Finally, nasals are lost before voiced stops, as in Modern Greek.
• Result at stage 4: word-final allophonic voicing.

b. Scenario 5: sound change plus analogy
Stage 1: saz atasa, saz dasa, sas tasa (final voicing assimilation)
Stage 2: saz tasa saz dasa sas tasa (aphaeresis)
Stage 3: saz tasa saz dasa saz tasa (analogical generalization of voicing)

At Stage 1, final obstruents undergo voicing assimilation. At Stage 2 voicing assimilation becomes opaque because initial vowels that trigger it are lost. Then the voiced obstruent is analogically generalized to all environments.
• Result at stage 3: word-final voicing

Kiparsky's Scenario 4 does indeed produce a system wherein stops only occur in word-final position if they are voiced. The post-nasal voicing process involved is extremely common cross-linguistically, leading one to think that this might be a promising candidate for producing a final voicing generalization. For the restriction of apocope to post-heavy position, however, the only typological parallel that Kiparsky mentions is a contested interpretation of Old English (see Minkova (1991) for a review of the relevant debate). The next step in the scenario is loss of nasals before (predictably voiced) stops including word-finally – a detail which seems auditorily peculiar since the cues of the nasal would be stronger than those of the stop in that position (cf. Wright (2004)). As a parallel, Kiparsky cites Modern Greek, where nasals are lost before voiced stops in some speech varieties in medial codas and across word boundaries (Arvaniti and Joseph (2000)), but it is unclear whether there are any examples with loss of nasals before predictably voiced word-final stops. More importantly, supposing that Scenario 4 is diachronically possible, its conditions seem narrow enough that our lack of any attestation of its outcome might be due to chance. We can hardly insist on the need for synchronic constraints to ban the outcome of a diachronic scenario which in and of itself does not seem very likely.

Scenario 5 also seems questionable because of the trajectory of analogy on which it depends. Consider the set of outcomes of this scenario in (24):

(24) expansion of Scenario 5 outcomes
- Stage 1: final voicing assimilation
  - /saz/: saz atasa, saz dasa, sas tasa
  - /sas/: sas atasa, saz dasa, sas tasa
- Stage 2: aphaeresis
  - /saz/: saz tasa, saz dasa, sas tasa
  - /sas/: sas tasa, saz dasa, sas tasa
- Stage 3: analogical generalization leaving only voiced obstruents finally
  - /saz/: saz tasa, saz dasa, saz tasa
  - /sas/: saz tasa, saz dasa, saz tasa

At stage 2, underlying word-final voiced obstruents surface as voiced except before an opaquely defined set of words with initial voiceless obstruents, while underlying word-final voiceless obstruents surface as voiceless except before initial voiced obstruents. Under such complex conditions, it is unclear whether learners would be at all likely to level the underlying contrast as needed to reach stage 3. Each of the candidate outcomes [sas] and [saz] is a minority allomorph for one or the other of the two underlying forms /sas/ and /saz/, so we question the likelihood of the forms becoming homophonous. If the problem were avoided by only levelling the allomorphs of original /saz/, this would leave voiceless surface codas in original *sas atasa* and *sas tasa* > [sas tasa], making it implausible for learners to postulate a final or coda voicing rule. Additionally, it is unclear why learners would proceed to stage 3 by levelling the underlying contrast in favour of the syntagmatically opaque and featurally marked option (final voiced obstruents); the voiceless alternative is globally

just as common, and at stage 2 its distribution is not opaque and thus plausibly an easier basis for generalization.

Thus, out of Kiparsky's five diachronic scenarios for final voicing, three (1-3) do not actually produce surface forms compatible with a final voicing analysis, and two (4-5) seem unlikely to do so.

Along with these considerations we can raise a broader concern with the theoretical framework Kiparsky deploys to rule out final voicing as a possible synchronic phonological process. His analysis crucially relies on the assumption that constraints can single out marked feature values but not unmarked feature values, with the consequence that "marked feature values [can be] suppressed [but not inserted] in 'weak' prosodic positions" (2006:222). Iverson and Salmons (2011) point out that in this scheme, phonologically *ex nihilo* feature insertion or addition is impossible. Both Iverson and Salmons (2011) and Vaux and Samuels (2005) provide numerous cases of coda enhancement effects that falsify this analysis.

To summarize what we have seen in this section, in identifying areas where phonetics in conjunction with diachronic change predict the emergence of marked systems which fail to materialize, de Lacy and Kiparsky have opened up an important empirical front in the debate concerning whether UG principles of markedness are needed. Of course, these arguments are only as valid as the empirical claims they make. While none of the languages that Blevins suggested as synchronic examples of final or postvocalic voicing incontrovertibly display such phonological processes, Middle Persian is a plausible candidate for post-voiced voicing of obstruents including most word-final obstruent tokens in the language – the closest thing to final voicing for which we have identified a plausible diachronic origin. Kiparsky's claim that the absence of final voicing points to synchronic constraints outside the natural influence of the phonetic apparatus on sound change is weakened by the fact that his own diachronic scenarios either fail to produce it or do so via typologically questionable paths.

In light of the above, it seems premature to conclude either that final voicing does occur or that it cannot. A continued and concerted effort must be made to uncover examples of the supposedly absent structures before any final conclusions can be reached, as Kiparsky rightly notes (2006: 234).

5. Conclusions

The final voicing phenomenon just discussed is ambiguous not solely with respect to whether or not it exists, but also as to whether or not its explanation would fall under the purview of UG. In the OT conception of UG currently employed by Kiparsky, whatever constraint(s) are responsible for the universal absence of final voicing would presumably fall under the rubric of CON, which constitutes the core of UG. If however final voicing is in fact impossible, as Kiparsky claims, the violable constraints of CON may prove insufficient to the task. Assume for example that we attempt to ban final voicing by employing the Generalized Alignment schema of McCarthy and Prince (1993) to formulate a constraint along the lines of ALIGN(PWORD R, [+stiff]), which would punish output candidates whose rightmost segment is voiced. A constraint such as this would prevent grammars wherein it was appropriately ranked from producing final voicing. By the OT tenet of Violability, though, such a constraint could also be outranked for example by a minimally different constraint ALIGN(PWORD R, [-stiff]), which if it also dominated the relevant faithfulness constraints would generate final voicing. In the absence of an explicit theory of the contents of CON, there is nothing to bar such a constraint and hence

there is no explanation for the putative universal lack of final voicing processes. In light of this fact, a metaconstraint on GEN or CON may be required to ensure the correct typological result, and the membership of such a metaconstraint in UG vs. the LAD is not entirely clear.

Problems of this sort in delineating the domains of UG and the LAD led us at the outset of this chapter to suggest a new division of labor between the terms UG and LAD. In this scheme, the territory covered by the term "UG" in most work in the field is split in two, with "UG" referring specifically to the initial state of the grammar that all normal humans begin with, and "LAD" referring to the range of entities that learners employ in tandem with PLD to generate a lexicon and transmute the initial state into a steady state grammar that can map entries from the lexicon onto the desired outputs, and vice versa. We saw that many theoretical constructs commonly attributed to UG, such as the set of distinctive features, may be more properly viewed as components of the LAD in this division of labor, though this depends on the particular phonological theory one adopts. We saw moreover that arguments over the existence of such components, and of UG in general, typically revolve around questions of their language-specificity, rather than assailing the existence of UG as empiricists and the media often claim.

As such questions do not actually get at the (non)existence or nature of UG, we then looked in closer detail at two case studies that promised to actually shed light on the matter. Our investigations of *NT and final voicing illustrated how for at least these sorts of cases, the arguments for phonological elements of UG submit at least as well to historical, phonetic, or other extra-UG explanations, raising the possibility that the phonological content of UG might be significantly sparser than most phonologists assume.

What really needs to be investigated at this point is the class of phonological phenomena that cannot be so readily accounted for without recourse to UG and/or the LAD, such as the spontaneous appearance in first, second, and toy language acquisition of phonological phenomena including derived environment effects, identity avoidance effects, and local ordering effects (cf. Vaux (2012) on Korean language games). Such phenomena provide a promising and largely unexplored area for future research into UG and the LAD, which moreover promises to bring phonological theory closer to its neighbor, experimental psychology, and may provide novel and unexpected insights into the phonological component of our linguistic endowment.

References

Anderson, Janet. 1983. The difficulties of English syllable structure for Chinese ESL learners. Language Learning and Communication 2:53-61.

Arvaniti, A. and Brian Joseph. 2000. Variation in voiced stop prenasalization in Greek. Glossologia 11-12:131-166.

Back, Michael. 1978. Die sassanidischen Staatsinschriften. Leiden: Brill.

Blevins, Juliette. 2004. Evolutionary Phonology. Oxford: Oxford University Press.

Blevins, Juliette. 2006a. A theoretical synopsis of Evolutionary Phonology. Theoretical Linguistics 32/2, 117-166.

Blevins, Juliette. 2006b. Reply to commentaries. Theoretical Linguistics 32/2, 245-256.

Blevins, Juliette and Andrew Garrett. 1998. The origins of consonant-vowel metathesis. Language 74:508-556.

Blust, Robert. 2004. Austronesian nasal substitution: A survey. Oceanic Linguistics 43.1:73-148.
Browman, Catherine and Louis Goldstein. (1990). Tiers in articulatory phonology. In J. Kingston & M. E. Beckman (Eds.), Papers in Laboratory Phonology I (pp. 341-376). Cambridge: Cambridge University Press.
Calabrese, Andrea. 1988. Towards a theory of phonological alphabets. Doctoral dissertation, MIT.
Chang, S., Plauché M. & John Ohala. 2001. Markedness and consonant confusion asymmetries. In: Hume E. & K. Johnson (eds). The Role of Speech Perception in Phonology. NY: Academic Press. 79-101.
Chomsky, Noam. 1968. Language and Mind. New York: Harcourt Brace Jovanovich.
Chomsky, Noam. 2007. Approaching UG from below. In Uli Sauerland and Hans Martin Gärtner, eds., Interfaces + recursion = language? Chomsky's Minimalism and the view from syntax-semantics, pp. 1-29. New York: Mouton de Gruyter.
Chomsky, Noam and Morris Halle. 1968. The sound pattern of English. New York: Random House.
Chomsky, Noam, Marc Hauser, and Tecumseh Fitch. 2002. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? Science 298: 1569-1579.
Clements, G. N. 1985. The geometry of phonological features. Phonology Yearbook 2:225-252.
Clements, G. N. 1987. Phonological feature representation and the description of intrusive stops. In A. Bosch, B. Need, and E. Schiller (eds.), Parasession on Autosegmental and Metrical Phonology, CLS 23.2:29-50. Chicago: Chicago Linguistic Society.
Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In John Kingston and Mary Beckman, eds., Papers in Laboratory Phonology I: Between the grammar and the physics of speech, 283-333. Cambridge: Cambridge University Press.
Clements, G. N. 2003. Feature economy in sound systems. Phonology 20:287–333.
Cowie, Fiona. 1999. What's Within? Nativism Reconsidered. New York: Oxford University Press.
Cristi, Alejandrina and Amanda Seidl. 2008. Is Infants' Learning of Sound Patterns Constrained by Phonological Features? Language Learning and Development 4.3:203-227.
Davidson, Lisa, Peter Jusczyk, and Paul Smolensky. 2004. The initial and final states: theoretical implications and experimental explorations of Richness of the Base. In Constraints in phonological acquisition, René Kager, Joe Pater, and Wim Zonneveld, eds., 321-368. Cambridge: Cambridge University Press.
De Lacy, Paul. 2002. The formal expression of markedness. Doctoral dissertation, University of Massachusetts-Amherst.
De Lacy, Paul. 2006a. Markedness: Reduction and preservation in phonology. Cambridge: Cambridge University Press.
De Lacy, Paul. 2006b. Transmissibility and the role of the phonological component. Theoretical Linguistics 32.2:185-196.
De Lacy, Paul and John Kingston. 2013. Synchronic explanation. Natural Language and Linguistic Theory 31.2:287-355.
Donegan, Patricia and David Stampe. 1979. The study of Natural Phonology. In Dinnsen, Dan (ed.). Current Approaches to Phonological Theory. Bloomington: Indiana University Press, 126-173.

Dresher, B. Elan. 2009. The Contrastive Hierarchy in Phonology. Cambridge: Cambridge University Press.
Eckman, Fred. 1981. On the naturalness of interlanguage phonological rules. Language Learning 31:195-216.
Eckman, Fred, Abdullah El Reyes and Gregory Iverson. 2001. Allophonic Splits in L2 Phonology: The Question of Learnability. Perspectives on Interlanguage Phonetics and Phonology. Monograph issue of International Journal of English Studies 1.1:21-51.
Edmondson, Jerold, John Esling, and Jimmy Harris. 2004. Supraglottal cavity shape, linguistic register, and other phonetic features of Somali. Manuscript, University of Texas at Arlington. http://ling.uta.edu/~jerry/somali.pdf.
Evans, Daniel Silvan (1852) Geiriadur Saesoneg a Chymraeg: An English and Welsh Dictionary, Denbigh: Thomas Gee.
Fanselow, Gisbert and Caroline Féry. 2002. Resolving conflicts in grammars: Optimality Theory in syntax, morphology, and phonology. Hamburg: Helmut Buske Verlag.
Fitzpatrick, Justin, Andrew Nevins, and Bert Vaux. 2004. Exchange rules and feature-value variables. Paper presented at the North American Phonology Conference. http://www.pauldelacy.net/polarity/fitzpatrick_nevins_vaux-2004-Phonological%20Polarity.pdf
Flemming, Edward. 2005. Deriving natural classes in phonology. Lingua 115:287-309.
Flynn, Darin. 2007. Too Many Solutions: Says Who?. Handout of paper presented at the Alberta Conference on Linguistics, Banff, Oct. 20. http://www.pec.ucalgary.ca/dflynn/files/dflynn/Flynn07b.pdf, retrieved 25 January 2012.
Fodor, Jerry. 2001. Doing without What's within: Fiona Cowie's critique of nativism. Mind 110.437:99-148.
Gabbard, Kevin. 2010. A phonological analysis of Somali and the guttural consonants. Bachelor's thesis, Ohio State University.
Gafos, Adamantios. 1998. A-templatic reduplication. Linguistic Inquiry 29: 515–527.
Goldsmith, John. 1976. Autosegmental phonology. Doctoral dissertation, MIT.
Hale, Mark and Charles Reiss. 2000. Phonology as cognition. In Phonological knowledge: Conceptual and empirical issues, Noel Burton-Roberts, Philip Carr, and Gerard Docherty, eds., 161-184. Oxford: Oxford University Press.
Hansson, Gunnar. 2001. Theoretical and Typological Issues in Consonant Harmony. Doctoral dissertation, University of California, Berkeley.
Hayes, Bruce. 1997. Phonetically driven phonology: The role of Optimality Theory and Inductive Grounding. Manuscript, UCLA. http://www.linguistics.ucla.edu/people/hayes/phonet/indgroun.pdf (retrieved May 2012).
Hayes, Bruce and Donca Steriade. 2004. Introduction: the phonetic bases of phonological markedness. In In Bruce Hayes, Robert Kirchner, and Donca Steriade (eds.), Phonetically-based phonology, pp. 1-32. Cambridge: Cambridge University Press.
Helgason, Petur and Catherine Ringen. 2008. Voicing and aspiration in Swedish stops. Journal of Phonetics 36:607-628.
Howe, Darin. 2004. Vocalic dorsality in Revised Articulator Theory. Manuscript, University of Calgary. http://www.ucalgary.ca/dflynn/files/dflynn/Howe2004.pdf (retrieved 29 July 2008).

Huffman, Marie and Thomas Hinnebusch. 1998. The phonetic structure of 'voiceless' nasals in Pokomo: implications for sound change. Journal of African Languages and Linguistics 1:1-20.
Inkelas, Sharon and Orhan Orgun. 1995. Level ordering and economy in the lexical phonology of Turkish. Language 71:763–93.
Itō, Junko. 1986. Syllable theory in Prosodic Phonology. Doctoral dissertation, University of Massachusetts-Amherst.
Iverson, Gregory and Ahrong Lee. 2006. Perception of Contrast in Korean Loanword Adaptation. Korean Linguistics 13:49-87.
Iverson, Gregory and Joseph Salmons. 2011. Final devoicing and final laryngeal neutralization. In The Blackwell Companion to Phonology, Marc van Oostendorp, Colin Ewen, Beth Hume, and Keren Rice, eds., 1622-1643. Oxford: Wiley-Blackwell.
Kager, René. 1999. Optimality Theory. Cambridge: Cambridge University Press.
Kang, Yoonjung. 2003. Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. Phonology 20. 219-273.
Kenstowicz, Michael. 1994. Phonology in generative grammar. Oxford: Blackwell.
Kenstowicz, Michael and Charles Kisseberth. 1979. Generative phonology: Description and theory. San Diego: Academic Press.
Kiparsky, Paul. 1973. Phonological representations. In Three Dimensions of Linguistic Theory (ed. O. Fujimura), pp. 1–135. Tokyo: TEC Corporation.
Kiparsky, Paul. 1985. Some Consequences of Lexical Phonology. Phonology Yearbook 2:82-138.
Kiparsky, Paul. 1993. Blocking in Nonderived Environments. In Sharon Hargus and Ellen Kaisse, eds., Studies in Lexical Phonology. San Diego: Academic Press.
Kiparsky, Paul. 2006. Amphichronic linguistics vs. Evolutionary Phonology. Theoretical Linguistics 32:217-236.
Kiparsky, Paul. 2008. Universals constrain change, change results in typological generalizations. In Jeff Good (ed.) *Linguistic Universals and Language Change*. Oxford, OUP.
Korn, Agnes. 2003. Balochi and the Concept of North-Western Iranian. In The Baloch and Their Neighbours: Ethnic and Linguistic Contact in Balochistan in Historical and Modern Times, Carina Jahani and Agnes Korn, eds., 49-60. Wiesbaden: Reichert.
Korn, Agnes. 2009. Lengthening of i and u in Persian. In Exegisti monumenta: Festschrift in honour of Nicholas Sims-Williams, Werner Sundermann, Almut Hintze, and François de Blois, eds., 197-213. Wiesbaden: Harrassowitz.
Koutsoudas, Andreas, Gerald Sanders, and Craig Noll. 1974. The application of phonological rules. Language 50.1:1-28.
Kuhl, Patricia and J. Miller. 1975. Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. Science 190.4209:69-72.
Leben, William. 1973. Suprasegmental phonology. Doctoral dissertation, MIT.
Lombardi, Linda. 2001. Why Place and Voice are different: Constraint-specific alternations in Optimality Theory. In Linda Lombardi (ed.), Segmental Phonology in Optimality Theory: Constraints and Representations, 13–45. Cambridge: Cambridge University Press.
Mackenzie, D. 1967. Notes on the transcription of Pahlavi. Bulletin of the School of Oriental and African Studies 30.1:17-29.
Mackenzie, D. 1971. A concise Pahlavi dictionary. London: Oxford University Press.

Major, Roy. 1987. A model for interlanguage phonology. In G. Ioup & S. Weinberger. (Eds.) Interlanguage Phonology (pp 101 - 124). Rowley, MA: Newbury House Publishers.
Major, Roy. 2001. Foreign accent: The ontogeny and phylogeny of second language phonology. Mahwah, NJ: Lawrence Erlbaum Associates.
McCarthy, John. 1988. Feature geometry and dependency: A review. Phonetica 45:84-108.
McCarthy, John. 2002. A Thematic Guide to Optimality Theory. Cambridge: Cambridge. University Press.
McCarthy, John. 2008. Doing Optimality Theory. Oxford: Blackwell.
McCarthy, John and Alan Prince. 1993. Generalized Alignment. Yearbook of Morphology 79-153.
McCarthy, John and Alan Prince. 1994. The emergence of the unmarked: Optimality in prosodic morphology. Proceedings of the North East Linguistics Society 24:333-379. Ed. Mercè Gonzàlez. Amherst, MA: GLSA.
Meillet, Antoine and J. Vendryes. 1968. Traité de grammaire comparée des langues classiques, 4th edition. Paris: Librairie Ancienne Honoré Champion.
Melchert, Craig. 1994. Anatolian historical phonology. Amsterdam: Rodopi.
Mielke, Jeff. 2004. The emergence of distinctive features. Doctoral dissertation, Ohio State University.
Minkova, Donka. 1991. The history of final vowels in English: The sound of muting. Berlin: Mouton de Gruyter.
Myers, Scott. 1997. Expressing phonetic naturalness in phonology. In Iggy Roca (ed.), Constraints and Derivations in Phonology, 125-152. Oxford: Oxford University Press, Oxford.
Myers, Scott. 2002. Gaps in factorial typology: the case of voicing in consonant clusters. Manuscript, University of Texas-Austin. Accessed online 14 May 2012 at http://roa.rutgers.edu/files/509-0302/509-0302-MYERS-0-0.PDF.
Nevins, Andrew and Bert Vaux. 2004. The transparency of contrastive segments in Sibe: evidence for Relativized Locality. Paper presented at GLOW 27, Thessaloniki, Greece.
Nevins, Andrew and Bert Vaux. 2008. Underlying representations that do not minimize grammatical violations. In Freedom of Analysis, Sylvia Blaho, Patrik Bye, and Martin Krämer, eds., pp. 35-61. Berlin: Mouton de Gruyter.
Nguyen, Thu and John Ingram. 2004. A corpus-based analysis of transfer effects and connected speech processes in Vietnamese English. Proceedings of the 10th Australian International Conference on Speech Science & Technology, Macquarie University.
Odden, David. 2005. Introducing phonology. Cambridge: Cambridge University Press.
Ohala, John. 1971. The role of physiological and acoustic models in explaining the direction of sound change. Project on Linguistic Analysis Reports (Berkeley) 15.25-40.
Ohala, John. 1972. How to represent natural sound patterns. Project on Linguistic Analysis (Berkeley) 16.40-57.
Ohala, John. 1975. Phonetic explanations for nasal sound patterns. In: C. A. Ferguson, Larry Hyman, & J. J. Ohala (eds.), Nasalfest: Papers from a symposium on nasals and nasalization. Stanford: Language Universals Project. 289 - 316.

Ohala, John. 1981. The listener as a source of sound change. In: C. S. Masek, R. A. Hendrick, & M. F. Miller (eds.), Papers from the Parasession on Language and Behavior. Chicago: Chicago Ling. Soc. 178 - 203.

Ohala, John. 1983. The origin of sound patterns in vocal tract constraints. In Peter MacNeilage (ed.), The production of speech. New York: Springer-Verlag. 189 - 216.

Ohala, John. 1997. Aerodynamics of phonology. In Proceedings of the Seoul International Conference on Linguistics, 20th edition, Seoul: Linguistic Society of Korea, 92-97, http://linguistics.berkeley.edu/~ohala/papers/SEOUL2-aero.pdf (retrieved 24 March 2012).

Ohala, John. 2005. Phonetic explanations for sound patterns. Implications for grammars of competence. In W. J. Hardcastle & J. M. Beck (eds.) A figure of speech. A festschrift for John Laver. London: Erlbaum. 23-38.

Ohala, John & J. Lorentz. 1977. The story of [w]: an exercise in the phonetic explanation for sound patterns. Berkeley Linguistic Society, Proceedings of the Annual Meeting 3:577-599.

Paster, Mary. 2012. Why doesn't phonology count? Talk presented at the 7th North American Phonology Conference (NAPhC), Concordia University, Montreal.

Pater, Joe. 1996. *NC̥. Proceedings of the North East Linguistics Society 26:227-239.

Paul, Ludwig. 2008. Kurdish language i: History of the Kurdish language. Encyclopaedia Iranica, retrieved on 26 January 2012 from http://www.iranicaonline.org/articles/kurdish-language-i.

Phillips, Colin, Tom Pellathy, and Alec Marantz. 2000. Phonological Feature Representations in Auditory Cortex. Manuscript, University of Maryland.

Pierrehumbert, Janet. 2001. Why phonological constraints are so coarse-grained. Language and Cognitive Processes 16:5-6, 691-698.

Pierrehumbert, Janet. 2006. The Statistical Basis of an Unnatural Alternation, in L. Goldstein, D.H. Whalen, and C. Best (eds), Laboratory Phonology VIII, Varieties of Phonological Competence. Mouton de Gruyter, Berlin, 81-107.

Pinker, Steven and Ray Jackendoff 2005. The faculty of language: what's special about it? Cognition 95.2:201-236.

Pisowicz, Andrzej. 1984. The development of the Middle Persian system of obstruents. In Middle Iranian Studies: Proceedings of the International Symposium Organized by the Katholieke Universiteit Leuven from the 17th to the 20th of May 1982, W. Skalmowski and A. Van Tongerloo, eds., 15-24. Leuven: Peeters.

Prince, Alan and Paul Smolensky. 1993/2004. Optimality Theory: Constraint interaction in generative grammar. RuCCS Technical Report 2, Rutgers University, Piscateway, NJ: Rutgers University Center for Cognitive Science. Revised version published 2004 by Blackwell, Oxford.

Pycha, Anne, Pawel Novak, Ryan Shosted, and Eurie Shin. 2003. Phonological rule-learning and its implications for a theory of vowel harmony. In Proceedings of WCCFL 22, Gina Garding and Mimu Tsujimura, eds., 423-435.

Reiss, Charles. 2003. Quantification in Structural Descriptions: Attested and Unattested Patterns. Linguistic Review 20:305–38.

Reiss, Charles. 2008. Constraining the Learning Path without Constraints, or The OCP and NoBanana. In Vaux and Nevins 2008.

Rendsburg, G. A. (1997) 'Ancient Hebrew phonology', in Alan S. Kaye (ed.) Phonologies of Asia and Africa, 65-83.

Rice, Keren. 1999. Featural markedness in phonology: Variation. GLOT 4(6,7):3-6.

Ringe, Donald. 2006. From Proto-Indo-European to Proto-Germanic. Oxford: Oxford University Press.
Rose, Sharon, and Rachel Walker. 2004. A typology of consonant agreement as correspondence. Language 80: 475–531.
Ross, Erica. 2011. Diachronic developments in the scope of phonological generalisations. Doctoral dissertation, University of Cambridge.
Samuels, Bridget. 2011. Phonological architecture: a biolinguistic perspective. Oxford: Oxford University Press.
Samuels, Bridget, Marc Hauser, and Cedric Boeckx. this volume. Do animals have Universal Grammar? A case study in phonology.
Schein, Barry and Donca Steriade. 1986. On geminates. Linguistic Inquiry 17:691-744.
Schmitt, Rüdiger. 1989. Altpersisch. In Corpus linguarum Iranicarum, Rüdiger Schmitt, ed., 56-85. Wiesbaden: Harrassowitz.
Schrijver, Peter. 1995. Studies in British Celtic Historical Phonology, Amsterdam: Rodopi.
Sihler, Andrew. 1995. New comparative grammar of Greek and Latin. Oxford: Oxford University Press.
Stampe, David. 1979. A dissertation on natural phonology. New York: Garland.
Sims-Williams, Nicholas. 1996. Eastern Iranian languages. Encyclopaedia Iranica vol. 7, pp. 649-652.
Steriade, Donca. 2008. The phonology of perceptibility effects: the P-map and its consequences for constraint organization. In The Nature of the Word: Essays in Honor of Paul Kiparsky, ed. K. Hanson and S. Inkelas, pp. 151-180. Cambridge, Mass.: MIT Press.
Strachan, John, Kuno Meyer, and Timothy Lewis. 1937. An introduction to Early Welsh. Manchester: Manchester University Press.
Tesar, Bruce and Paul Smolensky. 1998. Learnability in Optimality Theory. Linguistic Inquiry 29.2:229-268.
Thurneysen, Rudolf. 1946. A Grammar of Old Irish, revised edition, translated from the German by D. A. Binchy and Osborn Bergin; Dublin: Dublin Institute for Advanced Studies.
University of Wales, Trinity St. David. 2012. Geiriadur: Welsh-English/English-Welsh Online Dictionary, http://www.geiriadur.net/index.php? (retrieved April 4, 2012).
Vaux, Bert. 2003/2008. Why the phonological component must be serial and rule-based. Paper presented at the Annual Meeting of the Linguistic Society of America, Atlanta. Longer version published in 2008 in Rules, Constraints, and Phonological Phenomena, Bert Vaux and Andrew Nevins, eds., pp. 20-60. Oxford: Oxford University Press.
Vaux, Bert. 2003. Consonant epenthesis and the problem of unnatural phonology. Manuscript, Harvard University.
Vaux, Bert. 2005. Formal and empirical arguments for Morpheme Structure Constraints. Paper presented at the Annual Meeting of the Linguistic Society of America, Oakland.
Vaux, Bert and Andrew Nevins. 2007. Metalinguistic, shmetalinguistic: the phonology of shm-reduplication. CLS 39-1: The Main Session: Papers from the 39th Annual Meeting of the Chicago Linguistic Society, Jon Cihlar, Amy Franklin, David Kaiser, and Irene Kimbara, eds., pp. 702-721.

Vaux, Bert and Bridget Samuels. 2004. Explaining vowel systems: Dispersion Theory vs. natural selection. Paper presented at the Annual Meeting of the Linguistic Society of America, Boston. Extended draft available at http://cambridge.academia.edu/BertVaux/Papers/133548/Explaining_vowel_systems_Dispersion_Theory_vs._Natural_Selection.

Vaux, Bert and Bridget Samuels. 2005. Laryngeal markedness and aspiration. Phonology 22:395-436.

Vaux, Bert and Bridget Samuels. 2012. Consonant epenthesis and markedness. Manuscript, University of Cambridge.

Weber, Dieter. 2007. Pahlavi morphology. In Morphologies of Asia and Africa, Alan Kaye, ed., 941-973. Winona Lake, Indiana: Eisenbrauns.

Weiss, Michael. 2009. Outline of the Historical and Comparative Grammar of Latin. Ann Arbor: Beech Stave Press.

White, Lydia. 2003. Second Language Acquisition and Universal Grammar. Cambridge: Cambridge University Press.

Wood, Sidney. 1988. Vowel quantity and syllable structure in Welsh. Working Papers, Lund University Department of Linguistics, 33, 229-236, http://www.sciecom.org/ojs/index.php/LWPL/article/view/2625/2200 (retrieved April 4, 2012).

Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Bruce Hayes, Robert Kirchner, and Donca Steriade (eds.) Phonetically Based Phonology, Cambridge: Cambridge University Press, 34-57.

Xu, Zheng. 2004. The Interlanguage Phonology of Mandarin Learners of English and the Gradual Learning Algorithm. 9th Conference on Laboratory Phonology.

Yang, Charles. 2004. Universal Grammar, statistics, or both? TRENDS in Cognitives Sciences 8.10:451-456.

Yu, Alan. 2004. Explaining final obstruent voicing in Lezgian: Phonetics and history. Language. 80(1): 73-97.

Yu, Alan. 2011. Mergers and neutralization. In Marc van Oostendorp, Colin Ewen, Elizabeth Hume, and Keren Rice (eds.) The Blackwell Companion to Phonology, Oxford: Wiley-Blackwell.