A Minimalist Program for Phonology

Bridget Samuels
University of Maryland

To appear in *The Oxford Handbook of Linguistic Minimalism*

1 Introduction

Perhaps the single most important facet of Minimalism is that it turns the traditional, top-down approach to investigating Universal Grammar (UG) on its head. As Chomsky (2007:3) describes it, the Minimalist Program "seeks to approach the problem 'from bottom up': How little can be attributed to UG while still accounting for the variety of I-languages attained[...]?" This shift in perspective is particularly apparent in more recent Minimalist works (e.g., Chomsky 2004, 2005, 2007; Boeckx 2006, *inter alia*), but it is implicit in the Strong Minimalist Thesis, which dates back to the early 1990's. The Strong Minimalist Thesis is, as Boeckx (To appear) puts it, "a challenge to the linguistic community: Can it be shown that the computational system at the core of [the faculty of language] is optimally or perfectly designed to meet the demands on the systems of the mind/brain it interacts with?" We hope to push the Strong Minimalist Thesis as far as it can go because it encourages us to *make sense* of the language faculty's properties, not in isolation, but rather within the larger picture of cognition. What's more, pursuing this line of inquiry is bound to yield new understanding of the Conceptual-Intentional and Sensory-Motor systems, because it forces us to think about the legibility conditions imposed on the language faculty by those other modules.

Despite the way Minimalism has radically reshaped our view of the interfaces, it seems odd that, as van Oostendorp and van de Weijer (2005:3) remark, the Minimalist Program "has not been applied to phonology;" similarly, Pinker and Jackendoff (2005:220) state that "The Minimalist Program, in Chomsky's original conception, chooses to ignore... all the phenomena of

phonology." But there is no reason why this should be. The following quote summarizes one of the primary motivations behind the present work:

> "For decades, generative linguists have viewed the internal grammar in terms of the interplay of two types of factors: genetic endowment, generally referred to as Universal Grammar (UG), and experience—that is, exposure to e-language. In recent years this picture has been augmented by a third type of factor: general principles of biological/physical design. This new focus tends to worry those who had been hoping for a rich and articulate UG (see Pinker & Jackendoff [2005]), but on the other hand it is fully in line with minimalist thinking. A particularly welcome effect produced by this shift of focus is that we may now reassess the issue of formal similarities and dissimilarities between syntax and phonology. For many years, the dominant view has been that syntax and phonology are fundamentally different. [...] But general principles of design may very well be active in syntax and phonology in similar ways." (van Riemsdijk 2008:227)

Investigating phonology from this perspective lays the groundwork for testing and refining the arguments made by Bromberger and Halle (1989) in support of the view that phonology is fundamentally different from syntax (contra van der Hulst 2006 and Anderson 2006). Such work allows us to focus not on the question of *whether* phonology is different, but rather *how* it is different and *why* this is the case. If we are correct to emphasize the role of Third Factor principles in the architecture of grammar (Chomsky 2005), then this should be a fruitful endeavor.

I argue further that phonologists should take seriously the idea advanced in many recent Minimalist writings that phonology is an "ancillary" module, and that phonological systems are "doing the best they can to satisfy the problem they face: to map to the SM interface syntactic objects generated by computations that are 'well-designed' to satisfy C-I conditions" but unsuited to communicative purposes (Chomsky 2008:136). Phonology is on this view an afterthought, an externalization system applied to an already fully-functional internal language system. While some have taken this to suggest that phonology might be messy, and that we shouldn't expect to find evidence of 'good design' in it, there is another perspective which suggests instead that the opposite conclusion is warranted: phonology might be much

simpler (less domain-specific) than has previously been thought, making use of only abilities that already found applications in other cognitive domains at the time externalized language emerged (see §5; see also Mobbs 2008).

2 The Substance-Free Approach

I advocate here (and in recent work, most notably Samuels 2009a) for a bottom-up approach to phonology, made possible by treating the phonological module as a system of abstract symbolic computation, divorced from phonetic content. This approach has come to be called "substance-free phonology," as first described by Hale and Reiss (2000a,b). We argue that phonologists must stop the practice of "substance abuse," or misguidedly mixing the study of phonological form with the properties of phonetic content.[1] As summarized by Reiss (2008:258-259)

> "[Hale and Reiss (2000a,b)] conclude that the best way to gain an understanding of the computational system of phonology is to assume that the phonetic substance (say, the spectral properties of sound waves, or the physiology of articulation) that leads to the construction of phonological entities (say, feature matrices) *never* directly determines how the phonological entities are treated by the computational system. The computational system treats features as arbitrary symbols. What this means is that many of the so-called *phonological universals* (often discussed under the rubric of markedness) are in fact epiphenomena deriving from the interaction of extragrammatical factors like acoustic salience and the nature of language change. Phonology is not and should not be grounded in phonetics since the facts which phonetic grounding is meant to explain can be derived without reference to *phonology*."

The goal of substance-free phonology is to determine the nature of the universal core of formal properties that underlie all human phonological systems, regardless of the phonetic substance or indeed of the modality by which

[1] It is interesting to note that other cognitive scientists, such as Kaplan (1987 [1995]) and Pylyshyn (2003), also caution against "the seduction of substance" in their fields (computational linguistics and vision, respectively).

they are expressed. The substance-free approach is, like Minimalism, a research program rather than a theory; multiple different theories are being explored. All these theories share the following set of assumptions:

(1) The common basis of substance-free phonology (Blaho 2008:2)

- Phonology refers to the *symbolic computational system* governing the *signifiant*, i.e., the non-meaningful level of linguistic competence. Phonology is taken to be *universal* — common to all (natural human) languages and all modalities —, and *innate*. Phonological knowledge is part of UG, but phonetics is not.
- Phonological primes are substance-free, in that their phonetic interpretation is invisible to phonology, and thus does not play a role in phonological computation.
- Markedness and typological tendencies (in the sense of Greenberg 1957, 1978) are not part of phonological competence, but rather an epiphenomenon of how extra-phonological systems such as perception and articulation work.

One of the most salient arguments in favor of maintaining a substance-free phonology concerns the nature of what a theory of UG, and of phonological UG in particular, should seek to explain. Hale and Reiss (2008:3) set up the following hierarchy:

(2) Attested $\subset$ Attestable $\subset$ Humanly computable $\subset$ Statable

a. Attested: Cree-type grammars, English-type grammars, French-type grammars
b. Attestable: "Japanese" in 200 years, Joe's "English"
c. Humanly computable: p $\rightarrow$ s /__ r
d. Statable: V $\rightarrow$ V: in prime numbered syllables: paka$_2$nu$_3$tipa$_5$forse$_7$ $\rightarrow$ paka:nu:tipa:fose:

Clearly, the set of attested grammars is inappropriately small: it is, I hope, uncontroversial that the list of attested languages does not exhaust the possibilities provided by UG. Conversely, the set of statable languages is far too large: it seems like a pretty safe bet that no grammars refer to the set of prime numbers, or the sign of the Zodiac at the time of the utterance, or whether the interlocutor owns any blue collared shirts, etc.

The more pressing question is whether it is correct for a theory of UG to zero in on the set of attestable languages, or the humanly computable ones. Advocates of substance-free phonology hold that the idea "that every possible [grammar] should be instantiated by some attested language... is naïve, just as it is deeply naïve to expect that all logically possible permutations of genetic material in the human genome are actually attested in individual humans" (Vaux 2008:24). In Newmeyer's (2005) terms, we maintain that synchronic phonological theory should characterize only the set of *possible* languages, not *probable* ones. Instead, the biases typically attributed to formal markedness should be explained by reference to properties of our perception and production systems, and to sheer accidents of history; this shifts much of the burden to the theory of sound change.

If synchronic phonological theory's sole task is to describe what is a possible synchronic phonological pattern/process—in other words, if markedness is not part of phonological competence—then what accounts for the fact that some patterns and processes are more common than others? Blevins (2004:8-9; emphasis in the original) states the working hypothesis very clearly:

> "*[R]ecurrent synchronic sound patterns have their origins in recurrent phonetically motivated sound change.* As a result, there is no need to directly encode the frequent occurrence of these patterns in synchronic grammars themselves. Common instances of sound change give rise to commonly occurring sound patterns. Certain sound patterns are rare or unattested, because there is no common pathway of change which will result in their evolution."

The locus of explanation for phonological typology largely shifts to diachrony, because "many of the so-called *phonological universals* (often discussed under the rubric of markedness)" are not exceptionless, and "are in fact epiphenomena deriving from the interaction of extragrammatical factors such as acoustic salience and the nature of language change" (Hale and Reiss 2000a:167; emphasis in the original).[2] I will set the issue of diachrony

[2]Though one typical argument for dividing the labor between grammatical and extragrammatical explanation in this way is the fact that so-called phonological 'universals' typically have exceptions, I want to make clear that the presence of such exceptions is merely a *clue* that we should be looking to extragrammatical factors for an explanation of such tendencies; even exceptionless generalizations may not warrant grammatical explanations. As Hornstein and Boeckx (2009:81) write, when we turn our attention to true

aside here (but see Samuels 2009a, Chapter 2), and concentrate on the consequences of this view for the phonological portion of UG in the rest of the present work.

3 A Minimalist Program for Phonology

The work undertaken here opens the door for future research into questions which are independently raised and particularly timely given another consequence of the Minimalist Program (see Boeckx 2008): genuine variation within narrow syntax has been eliminated, being relegated instead to the lexicon and to morphology. As a result, there can be no more study of comparative *narrow* syntax, but careful investigation of phonological representations and processes can provide complementary data that is bound to inform our knowledge of syntax, both narrowly and broadly construed. Given what we now understand about syntactic architecture in light of Minimalism, such a study lays the groundwork for testing and refining the arguments made by Bromberger and Halle (1989) in support of the view that phonology is fundamentally different from syntax, a view which has been opposed by van der Hulst (2005) and Anderson (2006); we can begin to focus not on the question of *whether* phonology is different, but rather *how* it is different and *why* this is the case.

A major theme which I explore in recent work (especially Samuels 2008, 2009a) is that, while phonology and syntax may look similar on the surface—and this is not likely to be a coincidence—upon digging deeper, crucial differences between the two modules begin to emerge. One area where surface similarities hide striking differences is in the comparison between phonological syllables and syntactic phrases. Syllables and phrases have been equated by Levin (1985) and many others, with some going so far as to claim that phrase structure was exapted from syllable structure (Carstairs-McCarthy 1999). I argue these analogies are false, and that many of the properties commonly attributed to syllabic structure can be explained as well or better without positing innate structure supporting discrete syllables in the grammar. And in keeping with purely syntactic accounts of prosodic domains such

"I(nternalist)-Universals," or the laws of the faculty of language, as opposed to Greenbergian "E(xternalist)-Universals," "the mere fact that every language displayed some property P does not imply that P is a universal in the I-sense. Put more paradoxically, the fact that P holds universally does not imply that P is a universal."

as those proposed by Kahnemuyipour (2004), Ishihara (2007), and others, I move to eliminate the prosodic hierarchy as well, instead arguing that phonological phrasing is directly mapped from the phase structure of syntax, as we will discuss in the next section. This means phonological representations are free to contain much less structure than has traditionally been assumed, and in fact that they are fundamentally "flat" or "linearly hierarchical."

I posit only three basic computational operations for phonology (see Samuels 2009a):

- SEARCH provides a means by which two elements in a phonological string may establish a probe-goal relation. The SEARCH algorithm, formulated by Mailhot and Reiss (2007), formalizes the system of simultaneous rule application proposed in Chomsky and Halle (1968:344): "to apply a rule, the entire string is first scanned for segments that satisfy the environmental constraints of the rule. After all such segments have been identified in the string, the changes required by the rule are applied simultaneously."

- COPY takes a single feature value or bundle of feature values from the goal of a SEARCH application and copies these feature values (onto the probe of the SEARCH).

- DELETE removes an element from the derivation.

The relationship of these phonological operations to their apparent counterparts in (narrow) syntax remain to be investigated in more depth. However, one crucial difference between syntactic and phonological structure-building operations—the fact that phonology lacks Merge (combine α and β symmetrically) can already be noted; Samuels and Boeckx (2009) argue that this has multiple important consequences which are responsible for some of the ultimately distinct characters of these two systems. The counterpart of Merge at PF is what we call Concatenate (combine α and β asymmetrically), which is not a primitive operation, but rather is decomposable into SEARCH and COPY creating new precedence relationships between segments (see Raimy 2000 on precedence and Samuels To appear on this application of SEARCH and COPY). Whereas iterative applications of Concatenate yield a flat structure, iterative applications of Merge yield a nested hierarchical structure: syntactic structures must be flattened, whereas linear order is a primitive in phonology (Raimy 2000). Also, since phonology lacks Merge, it

also follows that it lacks movement, since movement is a subspecies of Merge (Internal Merge or Re-Merge; Chomsky 2004). Without the possibility of re-merging the same element, the notion of identity is extrinsic in phonology, unlike in syntax (see Raimy 2003).

Further investigation of all these issues is, in my view, central to extending Minimalism into the phonological domain. However, for reasons of space I must focus here on only one topic, namely how it is possible to construe the syntax/phonology interface in such a way that phonology need not build its own domains, but can merely operate over the strings it receives from the syntax directly. It is also worth noting here that the picture sketched below is quite far from the *whole* story concerning the syntax/phonology interface. The exact mechanisms underlying such PF operations as linearization, vocabulary insertion, and copy deletion remain to be uncovered, as does the ordering of these operations; the present work focuses on only one particular, seemingly late, stage of the mapping from syntax to phonology.

4 Phonological Derivation by Phase

Throughout the generative era, several cyclic models of phonology have been proposed. The first of these was introduced by Chomsky, Halle, and Lukoff (1956). The phonological cycle became a crucial component of Chomsky and Halle (1968) and was adopted in syntax by Chomsky (1957). In phonology, this concept was later implemented as the 'strict cycle' of Kean (1974) and Mascaró (1976). The tradition of Lexical Phonology (& Morphology) begun by Kiparsky (1982) and Mohanan (1982) developed the idea of cyclicity further, building on Pesetsky (1979).[3]

Recently, a new movement in phonological theory has emerged, attempting to combine the insights of Lexical Phonology with Distributed Morphology (Halle and Marantz 1993) and the phase-based theory of narrow syntax and the syntactic interfaces developed in Chomsky (2001, 2008; derivation by phase or DbP). The theory presented here, which I call phonological derivation by phase (PDbP), falls under this umbrella. It takes as a starting point the conceptual argument laid out in the foundational work by Marvin (2002:74): "If we think of levels in the lexicon as levels of syntactic attachment of affixes, we can actually say that Lexical Phonology suggests that

[3]For a proper introduction to Lexical Phonology, see the works cited above, the papers in Kaisse and Hargus (1993), and McMahon (2000).

phonological rules are limited by syntactic domains, possibly phases."

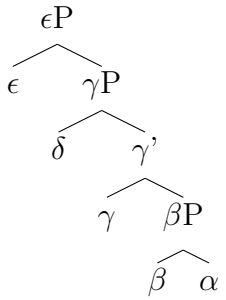
From a Minimalist standpoint, a model of grammar with synchronous cycles across the various modules is highly desirable. Indeed, it is this is the "best-case scenario" for computational efficiency according to Chomsky (2004:107). There is also a growing body of literature which argues that phases are required to regulate syntax's interfaces with the semantic and phonological components. See, for instance, Boeckx (2008) on how phases facilitate "wild-type" or "free" Merge and a conjunctivist semantics of the type proposed by Pietroski (2005 et seq.).

Moreover, PDbP also allows us to recognize the important contributions of cyclic models of phonology. For instance, all attempts to account for phonological opacity effects in a monostratal theory suffer from serious empirical or technical problems (see Vaux 2008 and references therein for discussion). Since the model proposed here relies on a cycle that is not proprietary to phonology, it is insulated from one family of recurring criticisms of Lexical Phonology, namely that its levels were poorly motivated and allowed to proliferate in an unconstrained manner (see, e.g., Itô and Mester 2003). In PDbP, by contrast, we expect evidence for the cycle to come from syntax and semantics in addition to (morpho)phonology. And there can be no *ad hoc* stipulation of cycles/levels if a phonological analysis must be responsible to, and grounded in, such external evidence; conversely, phonological phenomena should be able to provide evidence which is helpful for syntactic analysis.

4.1 Phases & Spell-Out

Before going any further, we should clarify how the basic phase architecture works. Consider a syntactic tree like the one below.

(3)



At certain points during the construction of this structure, the derivation is punctuated by the introduction of a phase head. What is crucial for present purposes is that phase heads initiate Transfer or Spell-Out, sending a chunk of the completed derivation to the semantic and phonological systems. Specifically, the complement of a phase head is the chunk that gets transferred, at the point when another phase head enters the derivation. Upon transfer, the "spell-out domain" (transferred chunk) is rendered opaque to further syntactic operations. This is formalized in the Phase Impenetrability Condition:

(4) Phase Impenetrability Condition (Chomsky 2001 version)
For [ZP Z ... [HP α [H YP]]]]: The domain of H is not accessible to operations at ZP, but only H and its edge.

Typically (or perhaps even necessarily; see Richards To appear), phase heads and non phase heads alternate with one another, so the chunks being transferred are larger than a single terminal. For (3) below, let us assume that only γ and ϵ are phase heads. The derivation will proceed as follows:

(5) a. Merge (β, α): α accessible to β.
b. Merge (γ, βP): β, α accessible to γ.
c. Merge (δ, γ'): γ accessible to δ.
d. Merge (ϵ, γP): δ, γ accessible to ϵ. βP transferred.

In the discussion to follow, I assume that Uriagereka's (1999) conception of Multiple Spell-Out (i.e., complex specifiers and adjuncts are spelled out alone) and Chomsky's phase framework can be simultaneously entertained. One means for accomplishing this is suggested by recent proposals such as Narita (2009), Narita and Samuels (2009), and Boeckx (2008), which argue that only simplex syntactic objects can undergo Merge: complex objects introduced on a left branch must therefore be reduced to simplex objects before they can be integrated with the main derivational spine. This is achieved by the transfer of all but the head of the mergee. That is to say, complex specifiers and adjuncts must be headed by phase heads.[4]

[4]Naturally, if we are going to pursue this type of theory, we must identify what is a phase head, and therefore what is a spell-out domain. Chomsky (2001 et seq.) takes C and transitive v to be phase heads; Legate (2003), Marvin (2002), Marantz (2008), and others argue that v must be a phase head in unaccusative and passive constructions

One important clarification is necessary in order to enable us to make broader use of the Phase Impenetrability Condition. In narrow syntax, "accessible to operations" essentially means eligible for movement (i.e., Internal Merge or Re-Merge), and able to participate in Agree. For phonological purposes, I will move forward under the assumption that an accessible string of phonology is visible to SEARCH and modifiable by COPY and DELETE. Now let us assume, then, that phase impenetrability holds in phonology, so each phonological string becomes inaccessible subsequent to the transfer of another string to the phonological component. By preventing reaching back too far into the derivation, the Phase Impenetrability Condition derives the effects previously attributed to the erasure of morpheme boundaries ("brackets") at the end of every cycle (Siegel 1974, Mohanan 1982), opacifying the results of earlier cycles. In other words, a rule can only affect something on its own cycle and/or the previous one, nothing more. The solution adopted here is similar in spirit to the Lexical Phonology tradition: word-building operations and phonological rules interleave, and the Phase Impenetrability Condition prevents modifying previous cycles after they are built.

Another idea which is crucial to PDbP is that phasal domains identifiable not just at the clausal level (i.e., *v*, C, etc.) but also within words. Parallel to *v*, Marantz (2001) establishes {*n, a*} as phase heads. In Distributed Morphology terms, following Marantz (1997), these elements are the categorial heads to which a-categorial roots must merge, and derivational affixes also belong to these classes. Marvin (2002) and Di Sciullo (2004, 2005) argue on multiple independent grounds that the Phase Impenetrability Condition holds for these 'morphological phases.' I argue in §5.2 of Samuels (2009a) that lexical rules are responsible to the Phase Impenetrability Condition on this smaller scale—a lexical rule has as its domain two adjacent morpheme-level Spell-Out domains—while post-lexical rules are responsible to the Phase Impenetrability Condition at the clausal level.

The strongest claim made by the PDbP approach is that spell-out domains are the *only* domains that phonology needs. In other words, both the levels of Lexical Phonology and the constituents of the prosodic hierarchy

as well. Crucially, T is not a phase head. Svenonius (2004), Bošković (2005), and Ott (2008), among others, argue for D as a phase head, and I will follow them here; I also follow Bošković (2005) in claiming that D need not be present in all languages or for all arguments. Other questions remain open, such as whether P is also a phase head (see Abels 2003). It is my hope that PDbP will open the door for phonological effects to shed some light on these unresolved matters.

come for free when we assume Distributed Morphology and a phasal syntax: phonological domains are directly imposed by morphosyntactic structure, and phonology need not erect any boundaries. It has been recognized for at least forty years (i.e., at least back to Chomsky and Halle 1968) that phonological domains correspond—in some fashion—to morphosyntactic ones. If the correspondence is not one of exact congruence, then phonology must construct (or adjust) boundaries. But if the correspondence *is* exact, then phonology can simply 'read' the structures it is given. Theories that assume exact correspondence subscribe to the 'direct reference' conception of the syntax/phonology interface; see Kaisse (1985), Odden (1990), and Cinque (1993). In recent literature, it is common to read that direct reference cannot be correct because there are apparent mismatches between syntactic and phonological domains. This is the position held by proponents of 'indirect reference' theories such as Selkirk (1984), Nespor and Vogel (1986), Truckenbrodt (1995), Seidl (2001), and many others. If PDbP is correct, there is no need to abandon direct reference for an indirect theory. In fact, the situation is even better: phonology doesn't have to read syntactic boundaries, it just applies to each chunk as it is received. PDbP can thus lead us to an understanding of phrase-level phonology that involves no boundary construction and eliminates the prosodic hierarchy.

4.2 Prosodic Hierarchy Theory

Since Selkirk (1978), and in classic works on prosodic hierarchy theory such as Selkirk (1984) and Nespor and Vogel (1986), a hierarchy of phonological constituents has been identified.[5] The most standard of these are (from smallest to largest, or weakest to strongest) the phonological word (ω), the phonological phrase (ϕ), the intonational phrase (I-phrase), and the utterance (U).

It is commonly (though not exceptionlessly) thought that this hierarchy of constituents obeys the conditions in (6)-(7) (Selkirk 1984, Nespor and Vogel 1986):

(6) Strict Layering Hypothesis
A given nonterminal unit of the prosodic hierarchy, X^P, is composed

[5]I will not seek to give a comprehensive primer in Prosodic Phonology/prosodic hierarchy theory here; I direct the reader to Inkelas and Zec (1995), on which I base the brief introductory discussion below, for an overview.

of one or more units of the immediately lower category, X^{P-1}.

(7) Proper Containment
A boundary at a particular level of the prosodic hierarchy implies all weaker boundaries.

That is, the prosodic hierarchy is non-recursive (though see Dobashi 2003 and Truckenbrodt 1995 et seq. for arguments to the contrary), and no levels can be skipped.[6]

The fundamental hypothesis of prosodic hierarchy theory is that the constituents that are suggested by these converging lines of evidence are correlated with, but not isomorphic to, syntactic constituents. For this reason, it is (proponents of the prosodic hierarchy claim) necessary to erect and adjust boundaries in the phonology, on the basis of syntactic information. Two general schools of thought have emerged on how this is construction is undertaken: the relation-based mapping approach represented by Nespor and Vogel (1986), and the edge- or end-based mapping approach represented by Selkirk (1986) and, in Optimality-Theoretic terms, Truckenbrodt (1995, 1999). I briefly summarize below how ϕ is constructed in each of these theories.

(8) Relation-based ϕ-construction (Nespor and Vogel 1986:168ff)
   a. *ϕ domain*
   The domain of ϕ consists of a C [clitic group] which contains a lexical head (X) and all Cs on its nonrecursive side up to the C that contains another head outside of the maximal projection of X.
   b. *ϕ construction*
   Join into an n-ary branching ϕ all Cs included in a string delimited by the definition of the domain of ϕ.
   c. *ϕ restructuring (optional)*
   A nonbranching ϕ which is the first complement of X on its recursive side is joined into the ϕ that contains X.

(9) End-based ϕ-construction (Truckenbrodt 1995:223)
A language ranks the two following universal constraints:

[6]The evidence for prosodic constituents falls into three major classes: (a) phonological rules for which they serve as domains of application, (b) phonological processes which occur at their edges (primarily suprasegmental, e.g., boundary tones), and (c) restrictions on syntactic elements relative to their edges (e.g., second position clitics).

a. ALIGN-XP, R: ALIGN (XP, R; ϕ, R)
For each XP there is a ϕ such that the right edge of XP coincides with the right edge of ϕ.

b. ALIGN-XP, L: ALIGN (XP, L; ϕ, L)
For each XP there is a ϕ such that the left edge of XP coincides with the left edge of ϕ.

Dobashi (2003) shows how the theories in (8) and (9) make different predictions with regards to the syntactic structure in (10):

(10) $[_{IP}\ \mathrm{NP}_{Subj}\ \mathrm{Infl}\ [_{VP}\ \mathrm{V}\ \mathrm{NP}_{Obj}\ ]\ ]$

The relation-based model in (8) will construct (11a), and if the optional restructuring rule applies, (11b). The end-based model in (9), if ALIGN-XP, R outranks ALIGN-XP, L, will construct only (11b).[7]

(11) ϕ boundaries for (10):

a. $(\mathrm{NP}_{Subj})_\phi\ (\mathrm{Infl\ V})_\phi\ (\mathrm{NP}_{Obj})_\phi$

b. $(\mathrm{NP}_{Subj})_\phi\ (\mathrm{Infl\ V\ NP}_{Obj})_\phi$

The two prosodic hierarchy models therefore agree on one thing, namely that the subject must always be phrased separately, which a great deal of literature on prosodic phrasing in SVO languages has shown is generally true (see Dobashi 2003, Chapter 2). However, they differ as to whether it is possible to phrase the object alone as well. The fact that prosodic hierarchy theory (in whatever its guise) predicts such a restricted set of prosodic constituents ("domain clustering") is often cited as an advantage. Inkelas and Zec (1995:548) write, "in making these predictions, the Prosodic Hierarchy Theory distinguishes itself dramatically from so-called direct access theories... in which each individual phonological rule may specify its own unique syntactic conditions. There is no expectation in such theories of any convergence or mutual constraining effect among rule domains." In the remainder of this section, I attempt to show that, while Inkelas and Zec's criticism may be valid for the particular direct reference theories formulated prior to 1995, within a phase-based model of grammar, direct reference is actually more

[7] An OT implementation of end-based ϕ construction, the Strict Layer Hypothesis, and Proper Containment requires many more constraints than just the ALIGN family, such as WRAP-XP, NONRECURSIVITY, EXHAUSTIVITY, LAYEREDNESS, & HEADEDNESS. See Truckenbrodt (2007) for an overview.

constrained and more accurate in its predictions than prosodic hierarchy theory, and far more parsimonious in its assumptions.

Another important difference between direct and indirect reference theories has to do with modularity. Paraphrasing Seidl (2000), both sides acknowledge that there are phonologically-relevant domains at the phrasal level; direct reference theories state these domains in terms of syntactic primes, while indirect theories state them in terms of phonological primes. This is not a matter of mere preference; adopting indirect reference violates the modular architecture of grammar. For indirect reference theories, prosodic constituents are constructed from a syntactic representation, as should be obvious from (8)-(9). And yet, for Optimality-Theoretic approaches which use constraints like the ones in (9),

> "prosodic structure is created by ALIGN and WRAP constraints *in the phonology*, i.e. the constraints at hand being interspersed with purely phonological constraints in the same constraint hierarchy. Mapping between morpho-syntax and phonology, which is what ALIGN and WRAP do, is a process that needs to be able to interpret morpho-syntactic structure — something that is impossible on modular grounds when sitting in phonology." (Scheer In press, fn. 14)

In short, if we want to maintain that phonological representations do not include syntactic information, then the indirect mapping approach is not viable (see also Scheer 2008b, especially §7.4).

Many other arguments against the prosodic hierarchy exist, particularly in light of Bare Phrase Structure (Chomsky 1995), in which it is impossible to refer to syntactic projections (i.e., XP), as both relation- and edge-based approaches must (see Dobashi 2003:10ff). I will not attempt to recap these arguments here. The analyses presented by Seidl (2000, 2001) and Scheer's (2008b) conceptual arguments are to my mind particularly devastating for prosodic hierarchy theory, and I encourage the reader to consult these works. I will limit myself to one very simple argument here: as I have already mentioned, the reason why indirect reference theories exist in the first place is that there are allegedly mismatches between syntactic structure and phonological domains. One famous mismatch, already noted by Chomsky and Halle (1968), is shown in (12). Brackets represent clause boundaries and parentheses represent I-phrase boundaries.

(12) a. Syntax: This is [the cat that caught [the rat that stole [the cheese]]]

b. Phonology: (This is the cat) (that caught the rat) (that stole the cheese)

However, a phase-based approach to syntax fares much better when it comes to approximating both syntactic and phonological phenomena. In fact, I believe it fares so well that there are no longer mismatches, and the argument for indirect reference therefore disappears. For example, one diagnostic for the phrasing in (12b) comes from stresses on *cat, rat*, and *cheese*. But these stresses can be generated by a rule which accents to the highest element in each clause-level Spell-Out domain (Kahnemuyipour 2004); the purported mismatch is therefore illusory.

4.3 Direct reference & phase domains

Ultimately, direct reference can be maintained only if a spell-out domain corresponds *always* and *exactly* to ϕ, and only if ϕ is the unique level of phonological domain necessary above the word level, can we have a truly direct reference theory of the interface. The purpose of the present section is to give proof of concept for a theory that, in pursuing this goal, eliminates the recourse to projections/labels that plague prosodic hierarchy theory. This pared-down syntax/phonology interface should be the null hypothesis, as has been stated most explicitly by Scheer (2008b), given that Spell-Out is precisely the operation that connects syntax and phonology.

It would be a massive undertaking to show that phase domains suffice for every rule with a claimed domain of ϕ, but I attempt to give proof of concept in Samuels (2009a, §5.6) by using vowel assimilation in Lekeito Basque, obstruent voicing in Korean, and second-position clitic placement in Sebro-Croatian as case studies. I demonstrate that the Phase Impenetrability Condition predicts the application and blocking contexts of both processes, both above and below the word level, in exactly the way I have just described. Further empirical studies in phase-based phonology include Seidl (2001) primarily on Bantu & Korean; Marvin (2002) on English & Slovenian; Kahnemuyipour (2004) on Persian, English, and German; Piggott and Newell (2006) and Newell (2008) primarily on Ojibwa; Sato (2006) on Taiwanese, French, Gilyak, Kinyambo, & Welsh; Ishihara (2007) on Japanese; Bachrach and Wagner (2007) on Portuguese; Michaels (2007) on Malayalam;

Kamali and Samuels (2008a,b) on Turkish; and more programmatically, Embick (2008) and Scheer (2008a,b). At present, however, I will restrict myself to describing how we can achieve the attested typology of phonological phrasing in SVO languages using a phase-based model.

Let us consider what prediction a phase-based system makes about ϕ-construction, compared to prosodic hierarchy theory. Dobashi (2003) argues that there are essentially four types of attested SVO languages, none of which exhibit $(\text{S V})_\phi$ $(\text{O})_\phi$ phrasing:

(13) Typology of ϕ-domains in SVO languages (Dobashi 2003:38)
   a. $(\text{S})_\phi$ $(\text{V})_\phi$ $(\text{O})_\phi$
   b. $(\text{S})_\phi$ $(\text{V})_\phi$ $(\text{O})_\phi$ or
      $(\text{S})_\phi$ $(\text{V O})_\phi$ if O is non-branching
   c. $(\text{S})_\phi$ $(\text{V O})_\phi$
   d. $(\text{S})_\phi$ $(\text{V O})_\phi$ or
      $(\text{S V})_\phi$ if S is non-branching

The type of language in (13a) is exemplified by French, as shown in (14), and the Aŋlɔ dialect of Ewe. Italian falls under (13b); see (15). Kimatuumbi (16) represents (13c), and Kinyambo (17) is of the type in (13d).[8] All the examples below are taken from Dobashi (2003, §2.2); I indicate the phonological phrasing with brackets.

[8]Dobashi takes only v and C to be phase heads. He notes that if the verb raises to v, as is commonly assumed, this will result in the subject and verb phrased together, and the object in its own ϕ domain. This prediction differs from that of prosodic hierarchy theory (recall (10)), and is undesirable from a typological perspective. Dobashi's answer to the mismatch between the *prima facie* predictions of the phase-based model and the typology in (13) is to modify the spell-out procedure. He argues that the leftmost element in a spell-out domain is actually not spelled out with the rest of the phase, but instead hangs back to establish the ordering of its phase with respect to the next phase, and only then is transferred. This has the effect of delaying spell-out of V (the leftmost element in v's complement domain) and the subject (the leftmost element in C's complement domain), resulting in the desired separate ϕ domains for the subject, verb, and object. This captures (13a) and part of (13b). Languages like (13c,d) are claimed to have V-to-T movement and object raising to Spec,vP such that nothing remains in the lowest spell-out domain; for this reason, the verb and object are phrased together. To account for the second options in (13b,d), Dobashi proposes a rule of restructuring which combines two ϕ-domains if one of them fails to meet the requirement that each ϕ minimally contains two prosodic words. Languages that exhibit the alternations in (13b,d) allow restructuring, while (13a,c) do not.

(14) [L' immigré] [envoyait] [un paquet] [à sa famille] (French)
the immigrant sent a package to his family
'The immigrant sent a package to his family'

(15) a. [Venderá] [questo leopardo] [in dicembre] (Italian)
sell.FUT.3SG this leopard in December
'He will sell this leopard in December'

b. [prenderá] [tordi] *or* [prenderá tordi]
catch.FUT.3SG thrushes
'He will catch thrushes'

(16) a. [Mamboondó] [aawíile] (Kimatuumbi)
Mamboondo die.PST.3SG
'Mamboondo died'

b. [naamwéeni nchéngowe Maliíya]
see.PST.1SG husband Mary's
'I saw Mary's husband'

(17) a. [abakozi bákajúna] (Kinyambo)
workers help.PST.3PL
'The workers helped'

b. [abakozi bakúru] [bákajúna]
workers mature help.PST.3PL
'The mature workers helped'

c. [okubon' ómuntu]
see person
'To see the person'

However, this typology omits crucial data concerning the phrasing of clitics in French, as Narita and Samuels (2009) discuss. A subject and/or object clitic is actually phrased together with the verb:

(18) [Donnez en] [à Marcel]
give it to Marcel
'Give it to Marcel'

(19) [Nous allons]
we go.1PL
'We go'

(20) [Marie le voit]
Marie him sees
'Marie sees him'

The same is (optionally) true of simplex objects in Aŋlɔ Ewe, as shown below:

(21) [mē] [kpɛ́ flɛ́-gé]
I stone buy-PRT
'I'm going to buy a stone'

Thus, Dobashi's typology should be amended, as it turns out there are only two distinct types of phrasing, not four; furthermore, French and Aŋlɔ Ewe do not in fact pattern identically, while French and Kinyambo do.[9]

(22) Revised typology of ϕ-domains
   a. $(\text{S})_\phi$ $(\text{V})_\phi$ $(\text{O})_\phi$
   $(\text{S})_\phi$ $(\text{V O})_\phi$/$(\text{O V})_\phi$ if O is non branching (Aŋlɔ Ewe, Italian, Kimatuumbi)
   b. $(\text{S V})_\phi$ $(\text{O})_\phi$ if S is non-branching
   $(\text{S})_\phi$ $(\text{V O})_\phi$/$(\text{O V})_\phi$ if O is non branching
   $(\text{S O V})_\phi$ if S and O are non-branching (French, Kinyambo)

Now we will see how to generate this typology in PDbP. We start with the basic assumption, as introduced earlier, that a phase head's complement domain (minus what has already been spelled out) is transfered to the phonology as a unit, corresponding to ϕ in prosodic hierarchy theory. We begin with the fairly standard assumption that the inventory of phase heads includes C (or at least one head in an expanded Left Periphery), v, and D. Say the verb raises to v or to T. It so happens that in all of the languages mentioned above, there is evidence for V-to-T raising, but even if the verb only raises to v, an *in situ* object will be spelled out by v separately from the verb.

We also have to account somehow for the typological fact that branching arguments behave differently from non-branching ones. In (13) and (22) we see this asymmetry in object phrasing, which Dobashi (2003) accounts for by saying that object raising is optional and can only be undergone by

[9]I ignore here the differences between languages with respect to allowing OV order with non-branching objects, since it is orthogonal to the issue at hand.

simplex objects. Narita and Samuels (2009) provide an alternative account under which movement of a branching object will still result in its being phrased separately, if movement of a branching phrase requires prior spell-out of the phrase head's complement. In either of these accounts, a simplex object which raises out of VP may be phrased with the verb, for instance if it instantiates D (e.g., a pronoun) or undergoes N-to-D raising (e.g., a proper name; see Longobardi 1994); there is also the possibility that some languages which do not exhibit overt determiners may not even have D, as argued by Bošković (2005); we predict that in such a language, a raised NP object could phrase with a verb in T or v regardless of branchingness.

Comparing Italian with Ewe and Kimatuumbi, we note that they are alike in having the subject undergo A'-movement to TopicP (on the connection with *pro*-drop in Italian, see Alexiadou and Anagnostopoulou 1998 and Frascarelli 2007). For example, a subject can be followed by a complementizer in Kimatuumbi, which shows that subjects can be quite high in this language. The fact that subjects and preposed elements all behave alike with regard to phonological rules diagnostic of the ϕ domain in this language also suggests that this may be correct. It has has been independently argued for Bantu that subjects typically appear in Topic (see Demuth and Mmusi 1997 and references therein on Sesotho and Bantu more generally; also Pak 2008 on Luganda). Simplex objects (at least; recall the previous paragraph) may raise in all three languages.

The difference between French and Kinyambo on the one hand and Italian, Ewe, and Kimatuumbi on the other hand is the height of the subject: in French, the subject has only moved as far as Spec,TP, so a non-branching subject can be phrased with a verb in T. In sum, the phonological phrasing exhibited in a particular language depends on the interaction of three things, holding constant the position of the verb in T/v: whether the subject is in an A' position, whether the object has raised out of VP, and whether the arguments branch. There is no need for any readjustment of domains on the way from syntax to phonology, or for any purely phonological conditions tying prosodic phrasing to branchingness.

If this purely phase-based account of phonological phrases is correct, it makes explaining domain span rules very simple: they are simply post-lexical rules that clause-level phase heads trigger on their complement domains. By tying spell-out domains directly to prosodic phrasing, we also derive the Maximal ϕ Condition of Richards (2004, 2006):

(23) Maximal ϕ Condition
A prosodic phrase ϕ (...ω, etc.) can be no larger than a phase.

The same works for domain limit rules, or phenomena that take place at the edges of ϕ-domains. As we have just established, the domains in question are the complement domains of clausal phase heads. The edge of this domain can be marked by a post-lexical rule carried by the phase head since, as I mentioned earlier (and see Samuels 2009a, §5.2), a post-lexical rule sees the entire clause-level Spell-Out domain as a single string without any internal boundaries.

The ultimate message which I hope to convey is that, if we want to understand cross-linguistic variation in phonology, we need to understand cross-linguistic variation in morphosyntax better. This calls for collaboration between phonologists, morphologists, and syntacticians, all working together towards the common goal of describing the range of linguistic structures that are available. This could shed light on several outstanding issues, such as the intriguing phonological differences between polysynthetic and less agglutinative languages. In the languages we have seen in this chapter, a clause-level phase defines a phonological phrase which may consist of several words (recall the Maximal ϕ Condition). This provides an interesting contrast with the conclusions of Compton and Pittman (2007), who argue that in Inuktitut, the phase defines a single prosodic word; Piggott and Newell (2006) argue the same for Ojibwa. This suggests that at the opposite end of the spectrum are isolating languages like Chinese: for them, it is almost as if every terminal defines a prosodic word. This could perhaps be thought of as the prosodic word being defined as a morpheme-level phase rather than a clause-level one.

Many details of PDbP remain to be negotiated, given that the syntax upon which a theory of the syntax/phonology interface must necessarily depend remains in flux. Nevertheless, already there is a quickly-growing list of empirical successes which have been achieved by tying phonological rule application directly to Spell-Out domains, and the phonological literature is rife with obvious candidates for phase-based analyses.

5 Phonology, Minimalism, & Evolution

To conclude, I would like to take a step back and consider the implications of this type of phonological theory for the Minimalist Program and for biolinguistic concerns such as how language may have evolved. This is an

important, though understudied, area in light of criticisms of the Minimalist Program which hinge on arguments made concerning phonology. Most notably, in their rebuttal of Hauser, Chomsky, and Fitch (2002), Pinker and Jackendoff (2005) take Chomsky to task for ignoring "all the phenomena in phonology" (p. 220) in the original Minimalist papers. But if the perspective for which I have advocated here is correct, then phonology need not be seen as a thorn in the side of the Minimalist Program.

In the previous sections we discussed some consequences of the Minimalist perspective for phonological theory and the syntax/phonology interface, but in what remains we will turn the tables and ask not what Minimalism can do for phonology, but what phonology can do for Minimalism. I believe it makes sense to frame this discussion in evolutionary terms: in short, the idea that the language faculty is simple and the idea that language evolved quickly and with a minimum of genetic changes go hand in hand. As Hornstein and Boeckx (2009:82) explain,

> "[I]n light of the extremely recent emergence of the language faculty, the most plausible approach is one that minimizes the role of the environment (read: the need for adaptation), by minimizing the structures that need to evolve, and by predefining the paths of adaptation, that is, by providing preadapted structures, ready to be recruited, or modified, or third factor design properties that emerge instantaneously, by the sheer force of physical laws."

Along these lines, in Samuels (2009b) and Chapter 6 of Samuels (2009a), I demonstrate on the basis of behavioral and physiological studies on animal cognition that all the cognitive abilities necessary for the phonological representations and operations argued for in the previous chapters are present in creatures other than *Homo sapiens* (even if not to the same degree) and in domains other than phonology or, indeed, language proper. This implies that nothing required by phonology is part of the faculty of language in the narrow sense (FLN, as opposed to the faculty of language in the broad sense, FLB), in the terms of Hauser, Chomsky and Fitch (2002) and Fitch, Hauser, and Chomsky (2005). In particular, the conclusion I draw from this investigation is that phonology may be entirely explainable through Third Factor principles pertaining to general cognition and the SM system (Chomsky 2005 et seq.).

This view accords with the evolutionary scenario developed by Hauser, Chomsky, and Fitch (2002) and Fitch, Hauser, and Chomsky (2005), who

view language not as something that evolved gradually as an adaptation for communication (cf. Pinker and Jackendoff 2005; Jackendoff and Pinker 2005), but rather emerged suddenly as a result of minimal genetic changes with far-reaching consequences. Particularly relevant is the distinction they make between the "Faculty of Language - Broad Sense" (FLB), including all the systems that are recruited for language but need not be unique to language, or to humans, an the "Faculty of Language - Narrow Sense" (FLN), which is the subset of FLB that is unique to our species and to language. At present, the leading hypothesis among proponents of this view is that FLN is very small, perhaps consisting only of some type of recursion (i.e., Merge) and the mappings from narrow syntax to the interfaces.

I therefore reject the claim made by Pinker and Jackendoff (2005:212) that "major characteristics of phonology are specific to language (or to language & music), [and] uniquely human," and their statement that "phonology represents a major counterexample" to the hypothesis proposed by Hauser, Chomsky, and Fitch (2002), namely that FLN consists of only recursion and the mapping from narrow syntax to the interfaces. What I suggest, in effect, is that the operations and representations which underlie phonology were exapted, or recruited from other cognitive domains for the purpose of externalizing language.[10]

Few authors have discussed phonology as it pertains to the FLN/FLB distinction. For example, Hauser, Chomsky, and Fitch (2002:1573) list a number of approaches to investigating a list of the S-M system's properties (shown below in (24)), and these are all taken to fall outside FLN. However, none of these pertain directly to phonological computation.

(24) a. Vocal imitation and invention
Tutoring studies of songbirds, analyses of vocal dialects in whales, spontaneous imitation of artificially created sounds in dolphins

b. Neurophysiology of action-perception systems
Studies assessing whether mirror neurons, which provide a core substrate for the action-perception system, may subserve gestural and (possibly) vocal imitation

c. Discriminating the sound patterns of language
Operant conditioning studies of the prototype magnet effect in

[10]On the possibility that language more generally is an exaptation, see among others Piattelli-Palmarini (1989), Uriagereka (1998), Boeckx and Piattelli-Palmarini (2005), Hauser, Chomsky, and Fitch (2002), and Fitch, Hauser, and Chomsky (2005).

macaques and starlings

d. Constraints imposed by vocal tract anatomy
Studies of vocal tract length and formant dispersion in birds and primates

e. Biomechanics of sound production
Studies of primate vocal production, including the role of mandibular oscillations

f. Modalities of language production and perception
Cross-modal perception and sign language in humans versus unimodal communication in animals

While all of these issues undoubtedly deserve attention, they address two areas—how auditory categories are learned, and how speech is produced—which are peripheral to the core of phonological computation. The most interesting two issues from my perspective are (c) and (f). These are of course very relevant to the debate over whether phonological features may be emergent and how phonological categories are learned; I discuss both issues in Samuels (2009a), Chapter 3. And the instinct to imitate, addressed in (a) and (b), is clearly necessary to language acquisition. However, investigating neither these nor any of the other items in (24) has the potential to address how phonological objects are represented or manipulated, particularly in light of the substance-free approach to phonology, which renders questions about the articulators (e.g., (d, e)) moot since their properties are totally incidental and invisible to the phonological system.

Two papers by Yip (2006a,b) outline a more directly relevant set of research aims. She suggests that, if we are to understand whether 'animal phonology' is possible, we should investigate whether other species are capable of the following:[11]

(25) a. Grouping by natural classes

b. Grouping sounds into syllables, feet, words, phrases

c. Calculating statistical distributions from transitional probabilities

d. Learning arbitrary patterns of distribution

e. Learning/producing rule-governed alternations

[11] Yip mentions two additional items which also appear on Hauser et al.'s list: categorical perception/perceptual magnet effects and accurate production of sounds (mimicry).

f. Computing identity (total, partial, adjacent, non-adjacent)

This list can be divided roughly into three parts (with some overlap between them): (25a,b) are concerned with how representations are organized, (25c,d) are concerned with how we arrive at generalizations about the representations, and (25e,f) are concerned with the operations that are used to manipulate the representations. I would add three more areas to investigate in non-linguistic domains and non-human animals:

(26) g. Exhibiting preferences for contrast/rhythmicity

h. Performing numerical calculations (parallel individuation and ratio comparison)

i. Using computational operations: search, copy, concatenate, delete

In Samuels (2009a,b), I present evidence that a wide range of animal species are capable of the tasks in (a-i), though it may be the case that there is no single species (except ours) in which all these abilities cluster in exactly this configuration. I show (contra Yip) that there is already a substantial amount of literature demonstrating this, and that it is reasonable to conclude on this basis that no part of phonology, as conceived in the present work, is part of FLN.

It is beyond the scope of this paper to review studies of animal cognition and behavior in any depth. However, let me briefly note that the evidence suggests that Pinker and Jackendoff's criticism of Hauser, Chomsky, and Fitch (2002) concerning phonology is unfounded, at least if the theory of phonological representations and operations proposed in Samuels (2009a) is close to correct. A wide range of animal species have been shown to group objects, extract patterns from sensory input, perform sequential objects, perform searches, engage in copying behaviors, and manipulate sets through concatenation. As conceived of here, phonology thus provides no challenge to the idea that FLN is very small, perhaps consisting of just recursion and the mappings from syntax to the Conceptual-Intentional and Sensory-Motor interfaces. Moreover, such a conclusion lends credence to the idea that the theory sketched in §3 is a biologically plausible model of phonological competence. The human phonological system is, in short, a domain-general solution to a domain-specific problem, namely the externalization of language. Far from posing a problem for the evolutionary scenario posed by Hauser et al., or for Minimalism, the evidence from phonology in fact supports the guiding hypotheses of both of these programs.