

# The Third Factor in Phonology

Bridget Samuels

This article attempts to investigate how much of phonology can be explained by properties of general cognition and the Sensorimotor system — in other words, third-factor principles, in support of the evolutionary scenario posed by Hauser *et al.* (2002a). It argues against Pinker & Jackendoff's (2005: 212) claim that "major characteristics of phonology are specific to language (or to language & music), [and] uniquely human," and their conclusion that "phonology represents a major counterexample to the recursion-only hypothesis." Contrary to the statements by Anderson (2004) and Yip (2006a, 2006b) to the effect that phonology has not been tested in animals, it is shown that virtually all the abilities that underlie phonological competence have been shown in other species.

*Keywords:* evolution of language; phonology; third factor

## 1. Introduction

The present work is a preliminary attempt to determine how much of human phonological computation (i.e., representations and operations) can be attributed to mechanisms which are present in other cognitive areas and in other species. In other words, I explore 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 [Sensorimotor system] interface syntactic objects generated by computations that are 'well-designed' to satisfy [Conceptual-Intentional system] 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 (e.g., Mobbs 2008) have taken this to suggest that phonology might be messy, and that we should not expect to find evidence of 'good design' in it, there is another perspective which suggests instead that the opposite conclusion is warranted: Even if the Conceptual-Intentional interface is more transparent than the Sensorimotor one, phonology

---

I wish to thank Cedric Boeckx, Marc Hauser, Ansgar Endress, Terje Lohndal, two anonymous reviewers, and audiences at the Harvard Mind, Brain & Behavior Initiative Colloquium and *BALE 2008* at York University for their helpful comments. All faults remain my own.



might nevertheless 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.

This view accords with the evolutionary scenario developed by Hauser *et al.* (2002a) and Fitch *et al.* (2005), who suggest that language may have emerged suddenly as a result of minimal genetic changes with far-reaching consequences (cf. Pinker & Jackendoff 2005 and Jackendoff & Pinker 2005, who see language as manifesting complex design).<sup>1</sup> Particularly relevant is the distinction that Hauser *et al.* (2002a) 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, and 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/or lexicalization<sup>2</sup> plus the mappings from narrow syntax to the interfaces. Pinker & Jackendoff (2005: 212) claim that phonology constitutes a problematic counterexample to this hypothesis because “major characteristics of phonology are specific to language (or to language & music), [and] uniquely human.” In this article, I investigate the extent to which Pinker & Jackendoff’s criticism is viable, first by examining what abilities animals have which are relevant to phonology, and then by sketching out an account which I develop more fully elsewhere (Samuels 2009a), which I argue is consistent with the view that FLN is quite limited.

## 2. What Does Phonology Require?

Few authors have discussed phonology as it pertains to the FLN/FLB distinction. For example, Hauser *et al.* (2002a: 1573) list a number of approaches to investigating the Sensorimotor system’s properties (shown below in (1)), and these are all taken to fall outside FLN. However, none of these pertain directly to phonological computation.

- (1) a. *vocal imitation and invention*

Tutoring studies of songbirds, analyses of vocal dialects in whales, spontaneous imitation of artificially created sounds in dolphins

<sup>1</sup> The relation of Hauser *et al.*’s claims to the Minimalist Program is somewhat controversial, and the authors themselves claim that the two are independent. At least from my personal perspective, they are two sides of the same coin.

<sup>2</sup> Hauser *et al.* focused on the idea that recursion might be the crucial component in FLN. However, it has proven difficult to pinpoint what is meant by recursion in the relevant sense, such that it may be unique to humans and to language. Another hypothesis to which I am sympathetic has been proposed by authors such as Spelke (2003) and Boeckx (in press). On their view, it is not recursion but rather lexicalization — the ability to embed any concept in a ‘lexical envelope’ which allows it to be recursively Merged — which arose uniquely in our species. For the purposes of the present inquiry, we may simply note that both of these hypotheses exclude phonology from FLN.

- 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 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 these are all issues which 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. Nevertheless, (1c) and (1f), which I discuss in Samuels (2009a: sect. 3.2.1), are particularly interesting. These are relevant to questions of phonological acquisition and the building of phonological categories, including the possibility that phonological features are emergent rather than innate (see Mielke 2008). And the instinct to imitate, addressed in (1a) and (1b), is clearly necessary to language acquisition. However, I leave these items out of the present discussion because neither these nor any of the other items in (1) have the potential to address how phonological objects are represented or manipulated, particularly in light of the substance-free approach to phonology I adopt (see Hale & Reiss 2000a, 2000b, 2008), which renders questions about the articulators (e.g., (1d–e)) moot since their properties are totally incidental and invisible to the phonological system.

Two papers by Yip (2006a, 2006b) 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:<sup>3</sup>

- (2) 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
- f. Computing identity (total, partial, adjacent, non-adjacent)

---

<sup>3</sup> 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).

This list can be divided roughly into three parts (with some overlap): (2a–b) are concerned with how representations are organized, (2c–d) are concerned with how we arrive at generalizations about the representations, and (2e–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 in other species:

- (2) g. Exhibiting preferences for contrast/rhythmicity
- h. Performing numerical calculations (parallel individuation and ratio comparison)
- i. Using computational operations: search, copy, concatenate, delete

In the sections to follow, I will present evidence that a wide range of animal species are capable of the tasks in (2a–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 — in other words, it may be that what underlies human phonology is a unique *combination* of abilities, but the individual abilities themselves may be found in many other species. 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 my ongoing work, is part of FLN. In section 3, I focus on the abilities which underlie (2a,b,h) — that is, how phonological material is grouped. Next, in section 4, I turn to (2c–g), or the ability to identify and produce patterns. Finally, in section 5, I discuss (2e,i), the abilities which have to do with symbolic computation.

Before turning to these tasks, though, I would like to address one major concern which might be expressed about the discussion to follow. This concern could be phrased as follows: how do we know that the animal abilities for which I provide evidence are truly comparable to the representations and operations found in human phonology, and what if these abilities are only analogous, not homologous? Admittedly, it is probably premature to answer these questions for most of the abilities we will be considering. But even if we discover that the traits under consideration are indeed analogous, all is not lost by any means. In connection with this, I would like to highlight the following statement from Hauser *et al.* (2002a: 1572):

Despite the crucial role of homology in comparative biology, homologous traits are not the only relevant source of evolutionary data. The convergent evolution of similar characters in two independent clades, termed ‘analogies’ or ‘homoplasies,’ can be equally revealing [(Gould 1976)]. The remarkably similar (but non-homologous) structures of human and octopus eyes reveal the stringent constraints placed by the laws of optics and the contingencies of development on an organ capable of focusing a sharp image onto a sheet of receptors. [...] Furthermore, the discovery that remarkably conservative genetic cascades underlie the development of such analogous structures provides important insights into the ways in which developmental mechanisms can channel evolution [(Gehring 1998)]. Thus, although potentially misleading for taxonomists, analogies provide critical data about adaptation under physical and developmental constraints. Casting the comparative net more broadly, therefore, will most likely reveal

larger regularities in evolution, helping to address the role of such constraints in the evolution of language.

In other words, analogs serve to highlight ‘third-factor’ principles — that is, general properties of biological/physical design (Chomsky 2005, 2007) — which might be at play, and help us to identify the set of constraints which are relevant to the evolutionary history of the processes under investigation. For example, both human infants and young songbirds undergo a babbling phase in the course of the development of their vocalizations. Even though we do not want to claim that the mechanisms responsible for babbling in the two clades are homologous, nevertheless:

[T]heir core components share a deeply conserved neural and developmental foundation: Most aspects of neurophysiology and development — including regulatory and structural genes, as well as neuron types and neurotransmitters — are shared among vertebrates. That such close parallels have evolved suggests the existence of important constraints on how vertebrate brains can acquire large vocabularies of complex, learned sounds. Such constraints may essentially force natural selection to come up with the same solution repeatedly when confronted with similar problems.

(Hauser *et al.* 2002a: 1572)

We may not know what those constraints are yet, but until we identify the homologies and analogies between the mechanisms which underlie human and animal cognition, we cannot even begin to tackle the interesting set of questions which arises regarding the constraints on cognitive evolution. The present study, then, provides a place for us to begin this investigation in the domain of human phonological computation. I also want to emphasize that the components of phonology in (1)–(2) are intended to be as theory-neutral as possible, though in section 6 I give a brief overview of Samuels (2009a), a theory which I argue is especially well-suited to Hauser *et al.*’s hypotheses regarding the evolution of language, and also congenial to the Minimalist conception of the architecture of grammar. Furthermore, the basic argument I present against Pinker & Jackendoff — namely, that phonology does not constitute a major problem for Hauser *et al.* or for the Minimalist Program — can certainly hold even if one does not adopt my particular view of phonology.

### 3. Grouping

Since the hypothesis put forward by Hauser *et al.* (2002a) takes recursion to be the central property of FLN (along with the mappings from narrow syntax to the conceptual-intentional and Sensorimotor interfaces), much attention has been paid to groupings, particularly recursive ones, in language. While phonology is widely considered to be free of recursion,<sup>4</sup> nevertheless grouping (of features, of

---

<sup>4</sup> Some authors have argued for recursion in the higher levels of the prosodic hierarchy (e.g., at the Prosodic Word level or above). See Truckenbrodt (1995) for a representative proposal concerning recursion at the Phonological Phrase level. Even if this is correct (though see Samuels 2009a: chap. 5), the recursive groupings in question are mapped from syntactic

segments, and of larger strings) is an integral part of phonology, and there is evidence that infants perform grouping or ‘chunking’ in non-linguistic domains as well; see Feigenson & Halberda (2004). Additionally, segmenting the speech stream into words or morphemes (or syllables) also depends on what is essentially the converse of grouping, namely edge detection. We will discuss edge detection and the extraction of other patterns in section 4.

Human beings are masters at grouping, and at making inductive generalizations. Cheney & Seyfarth (2007: 118) write that “the tendency to chunk is so pervasive that human subjects will work to discover an underlying rule even when the experimenter has — perversely — made sure there is none.” This holds true across the board, not just for linguistic patterns. With respect to other species, many studies beginning with Kuhl & Miller (1975) show that mammals (who largely share our auditory system) are sensitive to the many of the same acoustic parameters as define phonemic categories in human language (see further discussion in Samuels 2009a: sect. 3.2). Experiments of this type provide the most direct *comparanda* to the groupings found in phonology. Even from a substance-free perspective, such results are valuable because they shed light on the origins of biases in phonetic perception which give rise to phonological patterns (see Blevins 2004 and Samuels 2009a: chap. 2 for an explicit connection to the substance-free program).

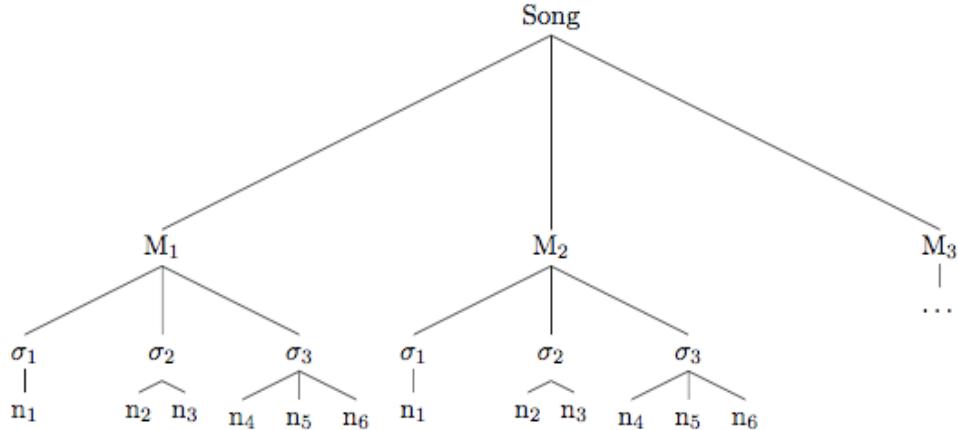
Also, relevantly to the processing of tone and prosody, we know that rhesus monkeys are sensitive to pitch classes — they, like us, treat a melody which is transposed by one or two octaves to be more similar to the original than one which is transposed by a different interval (Wright *et al.* 2000). They can also distinguish rising pitch contours from falling ones, which is an ability required to perceive pitch accent, lexical tone, and intonational patterns in human speech (Brosch *et al.* 2004). However, animals are generally more sensitive to absolute pitch than they are to relative pitch; the opposite is true for humans (see Patel 2008).

Another way of approaching the question of whether animals can group sensory stimuli in ways that are relevant to phonology is to see whether their own vocalizations contain internal structure. The organization of bird song is particularly clear, though it is not obvious exactly whether/how analogies to human language should be made. Yip (2006a) discusses how zebra finch songs are structured, building on work by Doupe & Kuhl (1999) and others. The songs of many passerine songbirds consist of a sequence of one to three notes (or ‘songemes’ as Coen (2006) calls them) arranged into a ‘syllable’. The syllables, which can be up to one second in length, are organized into motifs which Yip considers to be equivalent to prosodic words but others equate with phrases, and there are multiple motifs within a single song. The structure can be represented graphically as follows, where *M* stands for motif,  $\sigma$  stands for syllable, and *n* stands for note (modified from Yip 2006a):

---

structure, and are therefore not created by the phonological system alone. As an anonymous reviewer notes, this type of recursive structure is also quite different from the type found in syntax (for example, sentential embedding) which is limited in its depth only by performance factors.

(3)



There are a few important differences between this birdsong structure and those found in human phonology, some of which are not apparent from the diagram. First, as Yip points out, there is no evidence for binary branching in this structure, which suggests that the combinatory mechanism used by birds cannot be equated with binary Merge, but it could be more along the lines of adjunction or concatenation, which creates a flat structure; see section 6 and Samuels & Boeckx (2009). Second, the definition of a ‘syllable’ in birdsong is a series of notes/songemes bordered by silence (Williams & Staples 1992, Coen 2006). This is very unlike syllables, or indeed any other phonological categories, in human language. Third, the examples from numerous species in Slater (2000) show that the motif is typically a domain of repetition (as I have represented it above); the shape of a song is  $((a^x)(b^y)(c^z))^w$  with a string of syllables  $a, b, c$  repeated in order. This is quite reminiscent of reduplication. Payne (2000) shows that virtually the same can be said of humpback whale songs, which take the shape  $(a \dots n)^w$ , where the number of repeated components,  $n$ , can be up to around ten.

Both birdsong and whalesong structures are ‘flat’ (in the sense of Neeleman & van de Koot 2006) or ‘linearly hierarchical’ (in the sense of Cheney & Seyfarth 2007) — they have a depth of embedding which is limited to a one-dimensional string which has been delimited into groups, as in (4) — exactly what I argue in section 6 and in Samuels (2009a) for human phonology. It is interesting to note in conjunction with this observation that baboon social knowledge is of exactly this type, as Cheney & Seyfarth have described. Baboons within a single tribe (of up to about eighty individuals) obey a strict, transitive dominance hierarchy. But this hierarchy is divided by matrilines; individuals from a single matriline occupy adjacent spots in the hierarchy, with mothers, daughters, and sisters from the matriline next to one another. So an abstract representation of their linear dominance hierarchy would look something like this, with each  $x$  representing an individual and parentheses defining matrilines:

(4) (xxx)(xx)(xxxx)(xxx)(xxxxxx)(xxx)(x)(xxxx)

The difference between the baboon social hierarchy and birdsong, which I translate into this sort of notation below, is merely the repetition which creates a

motif (think of baboon individuals as corresponding to songemes and matri-lines corresponding to syllables):

$$(5) \quad \underbrace{(n_1)(n_2n_3)(n_4n_5n_6)}_{\text{motif}_1} \quad \underbrace{(n_1)(n_2n_3)(n_4n_5n_6)}_{\text{motif}_2}$$

There is evidence to suggest that, as in phonology (but strikingly unlike narrow syntax), the amount of hierarchy capable of being represented by animals is quite limited. In the wild, apes and monkeys very seldom spontaneously perform actions which are hierarchically structured with sub-goals and sub-routines, and this is true even when attempts are made to train them to do so. Byrne (2007) notes one notable exception, namely the food processing techniques of gorillas. Byrne provides a flow chart detailing a routine, complete with several decision points and optional steps, which mountain gorillas use to harvest and eat nettle leaves. This routine comprises a minimum of five steps, and Byrne reports that the routines used to process other foods are of similar complexity. Byrne further notes that “all genera of great apes acquire feeding skills that are flexible and have *syntax-like* organisation, with hierarchical structure. [...] Perhaps, then, the precursors of linguistic syntax should be sought in primate *manual* abilities rather than in their vocal skills” (Byrne 2007: 12; emphasis his). I concur that manual routines provide an interesting source of *comparanda* for the syntax of human language, broadly construed (i.e., including the syntax of phonology). Fujita (2007) has suggested along these lines the possibility that Merge evolved from an ‘action grammar’ of the type which would underlie apes’ foraging routines.

Other experiments suggest that non-human primates may be limited in the complexity of their routines in interesting ways. For example, Johnson-Pynn *et al.* (1999) used bonobos, capuchin monkeys, and chimpanzees in a study similar to one done on human children by Greenfield *et al.* (1972) (see also discussion of these two studies by Conway & Christiansen 2001). These experiments investigated how the subjects manipulated a set of three nesting cups (call them A, B, C in increasing order of size). The subjects’ actions were categorized as belonging to the ‘pairing,’ ‘pot,’ or ‘subassembly’ strategies, which exhibit varying degrees of embedding:<sup>5</sup>

- (6)    a. *Pairing strategy*: place cup B into cup C. Ignore cup A.
- b. *Pot strategy*: first, place cup B into cup C. Then place cup A into cup B.
- c. *Subassembly strategy*: first, place cup A into cup B. Then place cup B into cup C.

---

<sup>5</sup> The situation is actually substantially more complicated than this, because the subjects need not put the cups in the nesting order. To give a couple examples, putting cup A into cup C counts as the pairing strategy; putting cup A into cup C and then placing cup B on top counts as the pot strategy. I refer the reader to the original studies for explanations of each possible scenario. The differences between the strategies as I have described them in the main text suffice for present purposes.

The pairing strategy is the simplest, requiring only a single step. This was the predominant strategy for human children up to twelve months of age, and for all the other primates — but the capuchins required watching the human model play with the cups before they produced even this kind of combination. The pot strategy requires two steps, but it is simpler than the subassembly strategy in that the latter, but not the former, requires treating the combination of cups A + B as a unit in the second step. (We might consider the construction of the A + B unit as being parallel to how complex specifiers and adjuncts are composed ‘in a separate derivational workspace’ in the syntax; see Fujita 2007.) Human children use the pot strategy as early as eleven months (the youngest age tested) and begin to incorporate the subassembly strategy at about twenty months. In stark contrast, the non-human primates continued to prefer the pairing strategy, and when they stacked all three cups, they still relied on the pot strategy even though the experimenter demonstrated only the subassembly strategy for them. Though we should be careful not to discount the possibility that different experimental methodologies or the laboratory context is responsible for the non-humans’ performance, rather than genuine cognitive limitations, the results are consistent with the hypothesis that humans have the ability to represent deeper hierarchies than other primates. This is, of course, what we predict if only humans are endowed with the recursive engine that allows for infinite syntactic embedding (Hauser *et al.* 2002a).

Many other types of experimental studies have also been used to investigate how animals group objects. It is well known that a wide variety of animals, including rhesus monkeys, have the ability to perform comparisons of analog magnitude with small numbers (<4). They can discriminate between, for instance, groups of two and three objects, and pick the group with more objects in it. As Hauser *et al.* (2000) note, such tasks require the animal to group the objects into distinct sets, then compare the cardinality of those sets. Further data comes from Schusterman & Kastak (1993), who taught a California sea lion named Rio to associate arbitrary visual stimuli (cards with silhouettes of various objects printed on them). On the basis of being taught to select card B when presented with card A, and also to select card C when presented with card B, Rio transitively learned the A-C association.<sup>6</sup> Rio also made symmetric associations: when presented with B, she would select A, and so forth. We might consider these groups Rio learned to be akin to learning arbitrary pairings such as which phonemes participate in a given alternation (A and C bear the same relation to B), or in which contexts a particular process occurs (choose A in the context of B; choose B in the context of C).

The concept of ‘natural classes’ has also been studied in animals to a certain degree, though not in those terms. We can think of natural classes as multiple ways of grouping the same objects into sets according to their different properties (i.e., features). Alex the parrot had this skill: He could sort objects by color, shape, or material (reported by his trainer in Smith 1999). As regards the ability to group

---

<sup>6</sup> See also Addessi *et al.* (2008) on transitive symbolic representation in capuchin monkeys, and Cheney & Seyfarth (2007) on transitive inference involving social hierarchy in baboons. Cheney & Seyfarth also discuss both transitive social dominance and learning of symbolic representations in pinyon jays.

objects, then, I conclude that animals — especially birds and primates — are capable of the basic grouping abilities which phonology requires. They perceive (some) sounds categorically like we do; their vocalizations show linearly hierarchical groupings like ours; they can assign objects arbitrarily to sets like we do; they can categorize objects into overlapping sets according to different attributes like we do. Their main limitations seem to be in the area of higher-degree embedding, but this is (i) at best, a property of phonology which arises because of recursion in syntax, not from a recursive engine within phonology (or, if Samuels 2009a is correct to eliminate the prosodic hierarchy, not a property of phonology at all) and (ii) an expected result if, as Hauser *et al.* (2002a) hypothesize, recursion is a part of FLN and therefore not shared with other species.<sup>7</sup>

#### 4. Patterns

The next set of abilities we will consider are those which deal with extracting patterns from a data stream and/or learning arbitrary associations. As I mentioned in the previous section, I view pattern-detection as the flipside of grouping: A pattern is essentially a relation between multiple groups, or different objects within the same group. Thus, the ability to assign objects to a set or an equivalence class is a prerequisite for finding any patterns in which those objects participate; the abilities discussed in the previous section are very much relevant to this one as well.

Several experimental studies on animal cognition more generally bear on the issue of abstract pattern learning. One such study, undertaken by Hauser *et al.* (2002b), tested whether tamarins could extract simple patterns ('algebraic rules') like same–different–different (ABB) or same–same–different (AAB) from a speech stream. They performed an experiment very similar to one run on infants by Marcus *et al.* (1999). The auditory stimuli in both of these studies were of the form  $C_1V_1C_1V_1C_2V_2$  (the AAB condition) or  $C_1V_1C_2V_2C_2V_2$  (the ABB condition), such as *li-li-wi* or *le-we-we*. After habituating the infants/tamarins to one of these conditions, they tested them on two novel test items: one from the same class to which they had been habituated, and a second from the other class. The item with a different pattern than the habituated class should provoke a dishabituation response if the subjects succeed in learning the appropriate generalization based on the pattern in the stimuli presented during the training phase. Both infants and tamarins evidenced learning of these simple patterns; they were more likely to dishabituate to the item with the new pattern.

---

<sup>7</sup> A reviewer asks whether this implies animals have 'a little' recursion, and what that would even mean. I view the situation as an exact parallel to the difference between humans and animals in the domain of numerical cognition; perhaps the two dichotomies are indeed manifestations of the same cognitive difference, namely that only humans have a recursive engine (Merge), as suggested by Hauser *et al.* (2002a). While many animals (and young human children) seem to be able to represent small numerals, only suitably mature (and, perhaps, suitably linguistic) humans go on to learn the inductive principle, which allows them to count infinitely high. See discussion later in this section and section 5 for more discussion and references on numeracy in animals.

This type of pattern-extraction ability could serve phonology in several ways, such as the learning of phonological rules or phonotactic generalizations. Heinz (2007) showed that phonotactics (restrictions on the co-occurrence of segments, such as at the beginnings or ends of words) can be captured without any exceptions if three segments at a time are taken into account, so it seems on the basis of tamarins' success in the Hauser *et al.* experiment that learning phonotactics would not be out of their range of capabilities (though as we will soon see, tamarins may have independent problems with consonantal sounds that would interfere with this potential). Furthermore, phonotactics (and all attested phonological rules) can be modeled with finite-state grammars, as has been known since Johnson (1970). Here the somewhat controversial findings of Fitch & Hauser (2004) may also be relevant. At least under one interpretation of the data obtained by Fitch & Hauser, tamarins succeed at learning finite-state grammars but fail to learn more complicated phrase-structure grammars. If we accept these conclusions, then in theory — problems with consonants notwithstanding — we would expect that tamarins could learn any attested phonotactic restriction or phonological rule.

One of the most important obstacles facing a language learner/user falls into the category of pattern-extraction. This difficult task is parsing the continuous speech stream into discrete units (be they phrases, words, syllables, or segments). This speaks directly to (2b–c). Obviously, segmenting speech requires some mechanism for detecting the edges of these units. Since the 1950s, it has been recognized that one way to detect the edges of words is to track transitional probabilities, usually between syllables. If  $\text{Pr}(AB)$  is the probability of syllable B following syllable A, and  $P(A)$  is the frequency of A, then the transitional probability between A and B can be represented as:

$$(7) \quad TP(A \rightarrow B) = \frac{\text{Pr}(AB)}{\text{Pr}(A)}$$

The transitional probabilities within words are typically greater than those across word boundaries, so the task of finding word boundaries reduces to finding the local minima in the transitional probabilities. Numerous experimental studies suggest that infants do in fact utilize this strategy (among others) to help them parse the speech stream, and that statistical learning is not unique to the linguistic domain but is also utilized in other areas of cognition (see references in Gambell & Yang (2005)). With respect to the availability of this strategy in non-humans, Hauser *et al.* (2001) found that tamarins are able to segment a continuous stream of speech into three-syllable CVCVCV 'words' based solely on the transitional probabilities between the syllables. Rats are also sensitive to local minima in transitional probabilities (Toro *et al.* 2005).

While transitional probabilities between syllables are strictly local calculations (i.e., they involve adjacent units), some phonological (and syntactic) dependencies are non-adjacent. This is the case with vowel harmony, for instance, and is also relevant to languages with 'templatic' morphology, such as Arabic, in which a triconsonantal root is meshed with a different group of vowels depending on the part of speech which the root instantiates in a particular context. Comparing the results obtained by Newport & Aslin (2004) and

Newport *et al.* (2004) provides an extremely interesting contrast between human and tamarin learning of such patterns. Newport *et al.* tested adult humans and cotton-top tamarins on learning artificial languages, all with three-syllable CVCVCV words, involving the three different kinds of non-adjacent dependencies which I list below.

- (8) a. *Non-adjacent syllables*: the third syllable of each word was predictable on the basis of the first, but the second syllable varied.
- b. *Non-adjacent consonants*: The second and third consonants of each word were predictable on the basis of the first, but the vowels varied.
- c. *Non-adjacent vowels*: The second and third vowels of each word were predictable on the basis of the first, but the consonants varied.

Both humans and tamarins succeeded at learning the languages tested in the non-adjacent vowel condition. Humans also succeeded at the non-adjacent consonant condition. These results are expected, at least for the humans, because both of these types of dependencies are attested in natural language (in the guises of vowel harmony and templatic morphology, as already noted). Tamarins failed in the non-adjacent consonant condition, though this does not cast aspersions on the fact that they were able to learn non-adjacent dependencies; rather, it suggests that they have the cognitive capability needed to create the appropriate representations, but they might have difficulty distinguishing consonant sounds. In other words, their failure may not be due to the pattern-detection mechanism, but rather due to the input which was available to that mechanism. This interpretation is supported by the fact that tamarins succeeded at establishing dependencies between non-adjacent syllables.

From a phonological perspective, perhaps the most intriguing result is that humans failed at this non-adjacent syllable condition. Newport *et al.* (2004: 111) ask:

Why should non-adjacency — particularly syllable non-adjacency — be difficult for human listeners and relatively easy for tamarin monkeys? [...T]his is not likely to be because tamarins are in general more cognitively capable than adult humans. It must therefore be because human speech is processed in a different way by humans than by tamarins, and particularly in such a way that the computation of non-adjacent syllable regularities becomes more complex for human adults.

They go on to suggest that perhaps the syllable level is only indirectly accessible to humans because we primarily process speech in terms of segments (whereas tamarins process it in more holistic, longer chunks).<sup>8</sup> This is a possible contributor to the observed effect, but other explanations are available. I will propose one here.

What I would like to suggest is that, in effect, tamarins fail to exhibit a

---

<sup>8</sup> Alternatively, Newport *et al.* suggest, it could be that tamarins' shorter attention span reduces the amount of speech that they process at a given time; this would restrict their hypothesis space, making the detection of the syllable pattern easier. It is not obvious to me how this explains the tamarins' pattern of performance across tasks, however.

minimality effect.<sup>9</sup> Let us interpret the tamarins' performance in the non-adjacent consonant condition as suggesting, as I did above, that they either (for whatever reason) ignore or simply do not perceive consonants. Then for them, the non-adjacent syllable task differs minimally from the non-adjacent vowel task in that the former involves learning a pattern which skips the middle vowel. So rather than paying attention to co-occurrences between adjacent vowels, they have to look at co-occurrences between vowels which are one away from each other. It seems likely, as Newport *et al.* also suggest, that the adjacent vs. one-away difference represents only a small increase in cognitive demand. But for us, the non-adjacent syllable condition is crucially different — and this is true no matter whether we are actually paying attention to syllables, consonants, or vowels. These categories have no import for tamarins, but for humans, they are special. The dependency we seek in this condition is between two non-adjacent elements of the same category, which are separated by another instance of the same category. This is a classical minimality effect: if  $\alpha, \beta, \gamma$  are of the same category and  $\alpha > \beta > \gamma$  ( $>$  should be read for phonology as 'precedes' and for syntax, 'c-commands'), then no relationship between  $\alpha$  and  $\gamma$  may be established. This restriction is captured straightforwardly if the way linguistic dependencies are established (be that dependency an instance of Agree, harmony, or whatever else) is established by means of a search procedure which scans from  $\alpha$  segment by segment until it finds another instance of the same type (i.e.,  $\beta$ ), then stops and proceeds no further. If I am on the right track, then perhaps tamarins succeed where humans fail because their search mechanism does not work this way — which would be odd if minimality/locality restrictions arise from third-factor principles such as efficiency of computation — or more likely, that they do not represent the portions of the stimuli which they track as all belonging to the same abstract category of 'vowel' which is sufficient to trigger minimality effects for us.

A variety of other studies on primate cognition focus on the ability to learn sequences. Given that sequencing or precedence relationships are extremely important to language, particularly given the Minimalist emphasis on Merge in syntax and my parallel emphasis on concatenate in phonology, these studies are quite intriguing from a linguist's perspective. One apparent cognitive limitation of non-human primates relative to our species in the domain of pattern-learning is that they have extreme difficulty with non-monotonic sequences. Conway & Christiansen (2001) report on a number of studies which compare primates' performances on this kind of task. When presented with an 'artificial fruit' requiring four arbitrary actions to open it and thereby reveal a treat, chimpanzees and human preschoolers perform similarly; both succeed at learning the sequence.

However, another study highlights what seems to be a difference in the way humans and other primates plan and perform sequential actions. One

---

<sup>9</sup> Such effects have been discussed in terms of Relativized Minimality (Rizzi 1990) or the Minimal Link Condition (Chomsky 2000, 2004) in syntax and the No Line-Crossing Constraint (Goldsmith 1976) in auto-segmental phonology. I argue minimality in phonology and syntax emerges from the same underlying cause: A directional search mechanism which traverses strings of segments (see Mailhot & Reiss 2007, Samuels 2009a).

experiment undertaken by Ohshima (1997) tested human adults, Japanese monkeys, and a chimpanzee on the ability to learn an arbitrary pattern: They were presented with a touch screen with four different-sized colored circles on it and had to touch each one in sequence to receive a reward; the circles disappeared when touched. All the species succeeded in learning a monotonic pattern: touch the circles in order from smallest to largest or largest to smallest. They also all succeeded, but were slower, at learning non-monotonic patterns.<sup>10</sup> But as we will discuss in section 5, measurements of reaction times suggest the humans and monkeys used different strategies in planning which circles to touch.

Rhythm, too, is a type of pattern. Rhythmicity, cyclicity, and contrast are pervasive properties of language, particularly in phonology. Everything that has been attributed to the Obligatory Contour Principle (Leben 1973) fits into this category. Walter (2007) argues that these effects should be described not with a constraint against repetition (see also Reiss 2008), but as emerging from two major physical limitations: the difficulty of repeating a particular gesture in rapid succession, and the difficulty of perceiving similar sounds (or other sensory stimuli) distinctly in rapid succession. These are both extremely general properties of articulatory and perceptual systems which we have no reason to expect would be unique to language or to humans.

To date, perhaps the most direct cross-species test of the perception of human speech rhythm (prosody) comes from Ramus *et al.* (2000). In Ramus *et al.*'s experiment, human infants and cotton-top tamarins were tested on their ability to discriminate between Dutch and Japanese sentences under a number of conditions: one in which the sentences were played forward, one in which the sentences were played backward, and one in which the sentences were synthesized such that the phonemic inventory in each language was reduced to /s a l t n j/. The results of these experiments showed that both tamarins and human newborns were able to discriminate between these two unfamiliar and prosodically different languages in the forward-speech condition, but not in the backward-speech condition. A generous interpretation of these results would suggest "at least some aspects of human speech perception may have built upon preexisting sensitivities of the primate auditory system" (Ramus *et al.* 2000: 351). However, Werker & Voloumanos (2000) caution that we cannot conclude much about the processing mechanisms which serve these discrimination abilities; this is of particular concern given that the tamarins' ability to tell Dutch and Japanese apart was reduced in the reduced phonemic inventory condition. This may indicate that tamarins rely more strongly on phonetic cues rather than prosodic ones. Given the apparent importance of prosody for syntactic acquisition in human children — specifically, babies seem to use prosodic information to help them set the head parameter — Kitahara (2003: 38) puts forth the idea that

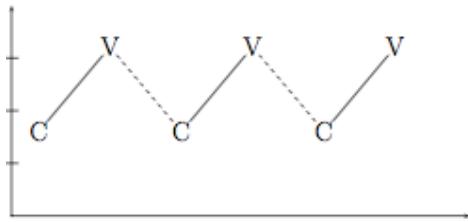
---

<sup>10</sup> In some situations, non-human primates fail entirely at learning non-monotonic patterns. For example, Brannon & Terrace (1998, 2000) found that while rhesus macaques taught the first four steps in a monotonic pattern could spontaneously generalize to later steps, they failed to learn a four-member non-monotonic pattern even with extensive training. It is not clear what to attribute the worse performance in the Brannon & Terrace studies to; there are too many differences between the paradigm they used and the one reported in the main text, including the species tested.

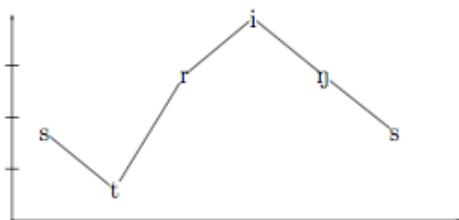
"cotton-top tamarins fail to discriminate languages on the basis of their prosody alone, because syntactic resources that require such prosodic-sensitive system [sic] might not have evolved for them." Though it is unclear how one might either support or disprove such a hypothesis, it is at the very least interesting to consider what prosody might mean for an animal which does not have the syntactic representations from which prosodic representations are built.

Another example of rhythmicity in speech is the wavelike sonority profile of our utterances, which is typically discussed in terms of syllable organization. Syllables range widely in shape across languages. In (9)–(10) I give examples from opposite ends of the spectrum: a series of three CV syllables in (9), and a syllable in (10) that has a branching onset as well as a coda, and additionally appendices on both ends. The relative heights of the segments in (9)–(10) represent an abstract scale of sonority (making no claim about the units of this scale).<sup>11</sup>

(9)

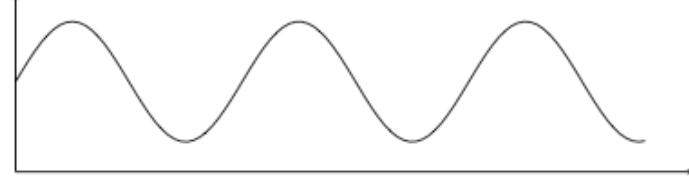


(10)



All syllables, from CV (9) to CCCVCC (10), combine to yield a sonority profile roughly as in (11):

(11)



The peaks and troughs may not be so evenly dispersed, and they may not all be of the same amplitudes, but the general shape is the same no matter whether the sonority values being plotted come from syllables that are CV, CVC, sCRV:CRs, and so forth, or any combination of these. This is hardly a new observation; it is over a century old (e.g., Lepsius & Whitney 1865, de Saussure

<sup>11</sup> I remain agnostic about the exact nature of sonority. However, see (among others) Ohala (1992) and Ohala & Kawasaki-Fukumori (1997) for arguments that it is a derived notion rather than a primitive one.

1916). Ohala & Kawasaki–Fukumori (1997: 356) point out that it is inevitable:

Just by virtue of seeking detectable changes in the acoustic signal one would create as an epiphenomenon, i.e., automatically, a sequence showing local maxima and minima in vocal tract opening or loudness. In a similar way one could find ‘peaks’ (local maxima) in a string of random numbers as long as each succeeding number in the sequence was different from the preceding one.

I have suggested in previous work that the ability to break this wave up into periods (based partially on universal and partially on language-specific criteria) aids with the identification of word boundaries: they tend to fall at the local minima or maxima in the wave (Samuels 2009a: sect. 3.3). And as we saw earlier in this section, we already know that both human infants and tamarins are sensitive to local minima (of transitional probabilities) in speech, which I believe suggests that this is a legitimate possibility.<sup>12</sup>

Animals from a wide variety of clades show preferences for rhythmicity in their vocalizations and other behaviors as well, though it is important to note that our own (non-musical) speech has no regular beat; while language does have a rhythm, it is not a primitive (see discussion in Patel 2008). Yip (2006b) mentions that female crickets exhibit a preference for males who produce rhythmic calls, and Taylor *et al.* (2008) discovered that female frogs prefer rhythmic vocalizations as well. Rhythmic behaviors, or the ability to keep rhythm, appear to be widespread in the animal kingdom. Gibbons produce very rhythmic ‘great calls,’ and while Yip (2006b: 443) dismisses this, saying that “the illusion of rhythm is probably more related to breathing patterns than cognitive organization,” this should hardly disqualify the data. For example, the periodic modulation of sonority in our speech is closely connected to opening and closing cycle of the jaw (Redford 1999, Redford *et al.* 2001), and it is widely accepted that the gradual downtrend in pitch which human utterances exhibit has to do with our breathing patterns. So for humans, too, there is at least some purely physiological component; however, the fact that females of various species prefer rhythmic calls shows that at the very least, there is also a cognitive component to animals’ perception of rhythmicity.

There are also some animals which synchronize the rhythms produced by multiple individuals. For example, frogs, insects, and bonobos all synchronize their calls; some fireflies synchronize their flashing, and crabs synchronize their claw-waving (see Merker 2000 and references therein). However, while elephants can be taught to drum with better rhythmic regularity than human adults, they do not synchronize their drumming in an ensemble (Patel & Iversen 2006).

Finally, we should note that it is extremely common for animals to exhibit ‘rule-governed’ behavior in the wild, and in their communicative behavior in particular. Cheney & Seyfarth (2007) make the case that baboon vocalizations are

---

<sup>12</sup> In all of the studies on tamarins (and human infants) of which I am aware, the shape of syllables tested does not extend beyond CV. As a reviewer suggests, it would be most informative to see studies which test a variety of syllable shapes — but note that tamarins’ difficulties with perceiving consonant sounds, as discussed earlier with regards to the Newport *et al.* (2004) experiments, would likely confound such investigations.

rule-governed in that they are directional and dependent on social standing. That is, a baboon will make different vocalizations to a higher-ranked member of the group than she will to a lower-ranked member. By this same rubric, vervet monkey grunts and chimpanzee calls should also be considered rule-governed; a number of articles on species ranging from treefrogs to dolphins to chickadees in a recent special issue of the *Journal of Comparative Psychology* (August 2008, vol. 122.3) devoted to animal vocalizations further cement this point. And as we saw in the previous section, both bird and whale songs obey certain combinatorial rules — in other words, they have some kind of syntax (in the broad sense of the term). Here the distinction made by Anderson (2004) and suggested in earlier work by Peter Marler is useful: Plenty of animals have a ‘phonological’ syntax to their vocalizations, but only humans have a ‘semantic’ or ‘lexical’ syntax which is compositional and recursive in terms of its meaning. Again, this reiterates Hauser *et al.*’s view that what is special about human language is the mapping from syntax to the interfaces (and particularly the LF interface, as Chomsky emphasizes in recent writings; see, e.g., Chomsky 2004), not the externalization system.

## 5. Operations

The final set of abilities which we will discuss are those which pertain to the phonological operations for which I argue in Samuels (2009a): SEARCH and COPY.<sup>13</sup> While these operations enjoy an elevated status in my work, as we will see in the next section, it is important to note that *any* theory of phonology, or of language in general, will have to make use of these operations. For example, Hornstein (2001) argues that insertion of an element into a linguistic derivation is copying from the lexicon, and I would add that it is very difficult to see how this copying might be done without a prior search into the lexicon. So, in short, one may contest my view that search, copy, and delete are the *only* operations in phonology, but it should not be seen as controversial that they play *some* role within the module. I also discuss here a fourth operation, concatenation. By this I mean the ability to connect morphemes — a root and an affix, for example — in a manner that creates a linear structure, not the nested hierarchical structure of Merge.<sup>14</sup> This concatenation mechanism properly belongs to the syntax–phonology interface, but since it operates at a stage at which phonological material has already been added (see Idsardi & Raimy, *in press*), as we know since some affixes are sensitive to phonological properties such as the stress pattern of the stem, it is relevant to the present work.

<sup>13</sup> I have little to say about the third operation which I posit, DELETE, but nothing suggests to me that this should be considered a domain-specific or species-specific ability.

<sup>14</sup> 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 sub-species 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). Samuels & Boeckx (2009) discuss this issue in greater detail.

Searching is ubiquitous in animal and human cognition. It is an integral part of foraging and hunting for food, to take but one example. The Ohshima (1997) study of sequence-learning by monkeys, humans, and a chimpanzee is an excellent probe of searching abilities in primates because it shows that, while various species can perform the multiple sequential searches required to perform the experimental task (touching four symbols in an arbitrary order), they plan out the task in different ways. The humans were slow to touch the first circle but then touched the other three in rapid succession, as if they had planned the whole sequence before beginning their actions (the ‘collective search’ strategy). The monkeys, meanwhile, exhibited a gradual decrease in their reaction times. It was as if they planned only one step before executing it, then planned the next, and so forth (the ‘serial search’ strategy).

Perhaps most interestingly of all, the chimpanzee appeared to use the collective search strategy on monotonic patterns but the serial search strategy when the sequence was not monotonic. That chimpanzees employ collective searches is corroborated by the results of a similar experiment by Biro & Matsuzawa (1999). The chimp in this study, Ai, had extensive experience with numerals, and she was required to touch three numerals on a touch-screen in monotonic order. Again, her reaction times were consistently fast after the initial step. But when the locations of the two remaining numerals were changed after she touched the first one, her reactions slowed, as if she had initially planned all three steps but her preparation was foiled by the switch. It is not clear to me exactly what should be concluded from the disparity between humans, chimps, and monkeys, but notice that the SEARCH mechanism proposed by Mailhot & Reiss (2007) and extended by Samuels (2009a, 2009b) operates in a manner consistent with the collective search strategy: scan the search space to find all targets of the operation to be performed, and then perform the operation to all targets in one fell swoop.

A close parallel to the COPY operation in phonology, particularly the copying of a string of segments as in reduplication, would be the patterns found in bird and whale songs. As we saw in section 3, Slater (2000) shows that for many bird species, songs take the shape  $((a^x)(b^y)(c^z))^w$ : That is, a string of syllables  $a, b, c$ , each of them repeated, and then the whole string repeated. We also saw that whale songs are similarly structured (Payne 2000). With respect to the copying of a feature from one segment to another (as in assimilatory processes), the relevant ability might be transferring a representation from long-term memory to short-term memory: extracting a feature from a lexical representation and bringing it into the active phonological workspace. This seems like a pre-requisite for any task which involves the recall/use of memorized information, and perhaps can be seen as a virtual conceptual necessity arising from computational efficiency (a prime source of third-factor explanation; see Chomsky 2005, 2007).<sup>15</sup>

As I mentioned in the previous two sections, concatenation serves both the

---

<sup>15</sup> If we think of copying as including imitative or mimicking behaviors, then this, too, is a very common ability. However, as Hauser (1996) stresses, monkeys and apes are not very strong vocal learners, as opposed to songbirds and cetaceans, which are quite skilled in this area. Nevertheless, monkeys’ learning is facilitated by watching a demonstration (Cheney & Seyfarth 2007), and Arbib (2005) argues that chimpanzees have the capacity for simple imitation that monkeys lack; humans have the capacity for complex imitation chimps lack.

ability to group and the ability to perform sequential actions. Without the ability to assign objects to sets or combine multiple steps into a larger routine, neither of these are possible. We have already seen that bird and whale songs have the kind of sequential organization which is indicative of concatenated chunks, and primates can perform multi-step actions with sub-goals.

I would like to suggest that concatenation may underlie the ‘number sense’ common to humans and many other species as well (for an overview, see Dehaene 1997, Lakoff & Nuñez 2001, Devlin 2005). This is perhaps clearest in the case of parallel individuation/tracking, or the ability to represent in memory a small number of discrete objects (< 4; see Hauser *et al.* 2000 and references therein). Shettleworth (1998) provides an overview of animal abilities in this domain, which have been shown for species as diverse as parrots and rats.

The idea that there is a connection between parallel individuation and concatenation is suggested by the fact that the speed of recognizing the number of objects in a scene decreases with each additional object that is presented within the range of capability (Saltzman & Garner 1948). This leads me to suspect, along with Gelman & Gallistel (1978) (but *contra* Dehaene) that such tasks require paying attention to each object in the array separately, albeit briefly. Lakoff & Nuñez (2001) also discuss a number of studies showing that chimpanzees (most notably Ai, whom we met previously as the subject of Biro & Matsuzawa’s 1999 study), when given rigorous training over a long period of time, can engage in basic counting, addition, and subtraction of natural numbers up to about ten. These tasks clearly involve the assignment of (sometimes abstract symbolic) objects to sets, which is the fundamental basis of concatenation. Conversely, subtraction or removal of objects from a set could be seen as akin to the delete operation; the ability to subtract has also been shown in pigeons. This and a number of other studies showing that primates, rats, and birds can both count and add with a fair degree of precision are summarized in Gallistel & Gelman (2005).

## 6. Approaching Phonology from Below

Now that we have seen an overview of animal abilities which seem to be relevant to phonological computation, I would like to take the next step and briefly describe how we might pursue a theory of phonology which employs virtually nothing besides these abilities plus the input given to phonology by (morpho-) syntax; the theory is laid out in detail in Samuels (2009a). This work is consistent with the ‘bottom-up’ approach to linguistic theory which is being pursued in syntactic circles. While more and more structure has been attributed to UG over the years, with the goal of reducing language acquisition to a manageable parameter-setting task for a child learner (i.e., taming Plato’s Problem), this perspective has shifted with the advent of the Minimalist Program (Chomsky 1995; MP), and particularly in the recent Minimalist works, (e.g., Chomsky 2004, 2005, 2007, Boeckx 2006, *inter alia*). Rather than asking how much UG must include, Minimalists argue, we must now turn this question on its head:<sup>16</sup>

---

<sup>16</sup> In advocating for a slimmer UG, it may seem that Minimalists find their aims more aligned

Throughout the modern history of generative grammar, the problem of determining the character of [the faculty of language] has been approached ‘from top down’: How much must be attributed to UG to account for language acquisition? The MP 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 [...]? (Chomsky 2007: 3)

Such a bottom-up approach to phonology is made possible by treating the phonological module as a system of abstract symbolic computation, divorced from phonetic content, pursuing the research agenda laid out by Hale & Reiss (2000a, 2000b). Along with Hale & Reiss and other ‘substance-free’ phonologists, I seek to investigate the universal core of formal properties that underlie all human phonological systems, regardless of the phonetic substance or indeed of the modality by which they are expressed. A major theme which I explore in recent work (Samuels 2009a, Samuels & Boeckx 2009) 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. In Samuels (2009a: chap. 5) I move to eliminate the prosodic hierarchy as well, instead arguing that phonological phrasing is directly mapped from the phase structure of syntax (see also Kahnemuyipour 2004, Ishihara 2007). 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.’ Thus, the theory of phonology for which I argue has fewer groupings, and fewer chances for those groupings to exhibit recursion or hierarchy, than most other approaches. This is true at virtually every level, from the sub-segmental to the utterance: I posit no feature geometry; no sub-syllabic constituency; no bracketing of morphemes; no prosodic hierarchy. The illusion of hierarchy is created by the pervasive processes of chunking (recall section 3) and repeated concatenation (recall section 5):

(12)	Concatenation	Chunking
	$\overline{\quad} + (\quad)$	$(\quad)(\quad)(\quad)(\quad)$
	$\overline{\quad} + (\quad)$	
	$\overline{\quad} + (\quad)$	

---

with those of neo-behaviorists/empiricists than was the case during earlier investigations in Principles-and-Parameters, as one reviewer points out. However, it is important to keep in mind that the driving force behind Minimalism (and the present work specifically) is not to deny that there is innate language faculty, but rather to search for the deep organizing principles of language whether they be specific to that faculty or not, and to present a theory which is consistent with the best current understanding of human evolution.

Still, nobody can deny the role of grouping/chunking in phonology: features group into segments, segments belong to natural classes on the basis of their featural composition, and segments group into longer strings such as syllables, morphemes, and phonological phrases. Of these last three types of groups, only the first is a truly phonological concept, since on my view phonology is a passive recipient of morphemes (strictly speaking, morpheme-level Spell-Out domains, which often but not always correspond to a single morpheme) and the chunks which correspond to phonological phrases (determined by the Spell-Out of phases common to narrow syntax, LF, and PF).<sup>17</sup>

I posit only three basic computational operations for phonology, as mentioned in the previous section:

- (A) SEARCH provides a means by which two elements in a phonological string may establish a probe-goal relation. The search algorithm, adapted from Mailhot & Reiss (2007), formalizes the system of simultaneous rule application proposed in Chomsky & Halle (1968: 344): “[T]o 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.”
- (B) 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).
- (C) DELETE removes an element from the derivation.

If I am correct in positing such a spare set of phonological representations and operations, then the research presented in the previous sections of the present work strongly suggests that at least the rudiments of all of the abilities which underlie this minimalist theory of phonology are present in other animal species, and in domains outside of language: That is, phonology may belong entirely to FLB.

## 7. Conclusions

I argue that the studies of animal cognition and behavior which I have presented here provide evidence that Pinker & Jackendoff's (2005) criticism of Hauser *et al.* (2002a) concerning phonology is unfounded, particularly if the theory of phonological representations and operations proposed in Samuels (2009a) is on

---

<sup>17</sup> Note that the model I assume is recursive in the sense that there are two types of Spell-Out domain, the morpheme-level and the clause-level, with the potential for several morpheme-level domains within a single clause-level one. However, these domains come directly from the narrow syntax, which is totally compatible with Hauser *et al.*'s hypothesis that syntax is the source — but crucially not the exclusive domain — of all recursive structures, and that once syntax is available, the modules with which it interfaces may be subject to modification.

the right track. Most conservatively, we can say that — *contra* Anderson (2004) and Yip (2006a, 2006b), we *have* tested for the building blocks of phonology in a wide range of species and found that they can group objects, extract patterns from sensory input, perform sequential objects, perform searches, engage in copying behaviors, and manipulate sets through concatenation. And more speculatively, we might tentatively conclude that, looking at the data we currently have, phonology provides little challenge to the idea that FLN is very small, perhaps consisting of just recursion or lexicalization and the mappings from syntax to the Conceptual-Intentional and Sensorimotor interfaces. This is most plausible if phonology is as conceived of in Samuels (2009a). The human phonological system would be, on this view, a domain-general solution to a domain-specific problem, namely the externalization of language. However, much research remains to be done in each and every one of the domains which I have discussed here, and I hope that the present work will be taken as an invitation to delve deeper and ask the more sophisticated questions which arise once we identify the basic points of potential consonance and divergence between human and animal cognition as far as phonology is concerned.

Another one of Pinker & Jackendoff's (2005) qualms with Hauser *et al.* — that the latter implicitly reject the popular hypothesis that 'speech is special' — should also be viewed skeptically. I do not deny the wide range of studies showing that speech and non-speech doubly dissociate in a number of ways which should be familiar to all linguists, as evidenced by aphasias, amusias, Specific Language Impairment, Williams Syndrome, autism, studies of speech and non-speech perception, and so on. Pinker & Jackendoff (2005) provide numerous references pointing to this conclusion, as does Patel (2008) with regards to language and music specifically (in this area the state of the art is changing rapidly, and the presence of a language/music dissociation is still an open and interesting question). But on the other hand, there is also a great deal of literature which shows that many species' vocalizations are processed in a different way from non-conspecific calls, or from sounds which were not produced by animals. This is true of rhesus macaques, who exhibit different neural activity — in areas including the analogs of human speech centers — and lateralization in response to conspecific calls (Gil da Costa *et al.* 2004). Perhaps we should amend the 'speech is special' hypothesis: speech is special (to us), in just the same way that conspecific properties throughout the animal kingdom often are; but there is nothing special about the way human speech is externalized or perceived in and of itself.

As a final note, consider the following set of characteristics which Seyfarth *et al.* (2005) ascribe to baboon social knowledge: it is representational, discretely-valued, linear-ordered, rule-governed, open-ended, modality-independent, combinatoric or concatenative, propositional, and linearly hierarchical. With the arguable exception of propositoriality (though cf. Bromberger & Halle 2000 on phonemes as predicates), this describes phonology perfectly. How can we maintain in light of this that the core properties of phonological computation are unique to language or to us?

## References

- Addessi, Elsa, Alessandra Mancini, Lara Crescimbene, Camillo Padoa-Schiopp & Elisabetta Visalberghi. 2008. Preference transitivity & symbolic representation in capuchin monkeys (*Cebus apella*). *PLoS ONE* 3, E2414. [doi:10.1371/journal.pone.0002414]
- Anderson, Stephen R. 2004. *Dr. Dolittle's Delusion: Animals and the Uniqueness of Human Language*. New Haven: Yale University Press.
- Arbib, Michael A. 2005. From monkey-like action recognition to human language. *Behavioral & Brain Sciences* 28, 105–167.
- Biro, Dora & Tetsuro Matsuzawa. 1999. Numerical ordering in a chimpanzee (*Pan troglodytes*): Planning, executing, & monitoring. *Journal of Comparative Psychology* 113, 178–185.
- Boeckx, Cedric. 2006. *Linguistic Minimalism*. Oxford: Oxford University Press.
- Boeckx, Cedric. In press. Some reflections on Darwin's problem in the context of Cartesian biolinguistics. In Anna-Maria Di Sciullo & Calixto Aguero (eds.), *Biolinguistic Investigations*. Cambridge, MA: MIT Press.
- Brannon, Elizabeth M. & Herbert S. Terrace. 1998. Ordering of the numerosities 1 to 9 by monkeys. *Science* 282, 746–749.
- Brannon, Elizabeth M. & Herbert S. Terrace. 2000. Representation of the numerosities 1–9 by rhesus macaques (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes* 26, 31–49.
- Bromberger, Sylvain & Morris Halle. 2000. The ontology of phonology (revised). In Noel Burton-Roberts, Philip Carr & Gerald Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*, 19–37. Oxford: Oxford University Press.
- Brosch, Michael, Elena Selezneva, Cornelia Bucks & Henning Scheich. 2004. Macaque monkeys discriminate pitch relationships. *Cognition* 91, 259–272.
- Byrne, Richard W. 2007. Clues to the origin of the human mind from primate observational field data. *The Japanese Journal of Animal Psychology* 57, 1–14.
- Carstairs-McCarthy, Andrew. 1999. *The Origins of Complex Language: An Inquiry into the Evolutionary Beginnings of Sentences, Syllables, and Truth*. Oxford: Oxford University Press.
- Cheney, Dorothy L. & Robert M. Seyfarth. 2007. *Baboon Metaphysics*. Chicago: University of Chicago Press.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2004. Beyond explanatory adequacy. In Adriana Belletti (ed.), *Structures and Beyond: The Cartography of Syntactic Structures*, vol. 3, 104–131. Oxford: Oxford University Press.
- Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 35, 1–22.
- Chomsky, Noam. 2007. Approaching UG from below. In Uli Sauerland & Hans-Martin Gärtner (eds.), *Interfaces + Recursion = Language?*, 1–29. Berlin: Mouton de Gruyter.

- Chomsky, Noam. 2008. On phases. In Carlos Otero, Robert Freiden & Maria-Luisa Zubizarreta (eds.), *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, 133–166. Cambridge, MA: MIT Press.
- Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.
- Coen, Michael H. 2006. Multimodal dynamics: Self-supervised learning in perceptual and motor systems. Cambridge, MA: MIT dissertation.
- Conway, Christopher M. & Morten H. Christiansen. 2001. Sequential learning in non-human primates. *Trends in Cognitive Sciences* 5, 539–546.
- Dehaene, Stanislas. 1997. *The Number Sense: How the Mind Creates Mathematics*. Oxford: Oxford University Press.
- Devlin, Keith. 2005. *The Math Instinct*. New York: Thunder's Mouth Press.
- Doupe, Allison J. & Patricia K. Kuhl. 1999. Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience* 22, 567–631.
- Feigenson, Lisa & Justin Halberda. 2004. Infants chunk object arrays into sets of individuals. *Cognition* 91, 173–190.
- Fitch, W. Tecumseh & Marc D. Hauser. 2004. Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380.
- Fitch, W. Tecumseh, Marc D. Hauser & Noam Chomsky. 2005. The evolution of the language faculty: Clarifications and implications. *Cognition* 97, 179–210.
- Fujita, Koji. 2007. Facing the logical problem of language evolution. Review of Jenkins (2004), *Variation and Universals in Biolinguistics*. *English Linguistics* 24, 78–108.
- Gallistel, Charles R. & Rochel Gelman. 2005. Mathematical cognition. In Keith Holyoak & Robert Morrison (eds.), *The Cambridge Handbook of Thinking and Reasoning*, 559–588. Cambridge: Cambridge University Press.
- Gambell, Tim & Yang, Charles D. 2005. Word segmentation: Quick but not dirty. Ms., Yale University, New Haven, CT.
- Gehring, Walter J. 1998. *Master Control Genes in Development and Evolution*. New Haven, CT: Yale University Press.
- Gelman, Rochel & Charles R. Gallistel. 1978. *The Child's Understanding of Number*. Cambridge, MA: Harvard University Press.
- Gil da Costa, Ricardo, Allen Braun, Marco Lopes, Marc D. Hauser, Richard E. Carson, Peter Herscovitch & Alex Martin. 2004. Toward an evolutionary perspective on conceptual representation: Species-specific calls activate visual & affective processing systems in the macaque. *Proceedings of the National Academy of Sciences* 101, 17516–17521.
- Goldsmith, John A. 1976. Autosegmental phonology. Cambridge, MA: MIT dissertation.
- Gould, Stephen J. 1976. In defense of the analog: A commentary to N. Hotton. In R. Bruce Masterson, William Hodo & Harry J. Jerison (eds.), *Evolution, Brain and Behavior: Persistent Problems*, 175–179. New York: Wiley.
- Greenfield, Patricia M., Karen Nelson & Elliot Saltzman. 1972. The development of rulebound strategies for manipulating seriated nesting cups: A parallel between action & grammar. *Cognitive Psychology* 3, 291–310.
- Hale, Mark & Charles Reiss. 2000a. Phonology as cognition. In Noel Burton-Roberts, Philip Carr & Gerald Docherty (eds.), *Phonological Knowledge: Con-*

- ceptual and Empirical issues, 161–184. Oxford: Oxford University Press.
- Hale, Mark & Charles Reiss. 2000b. Substance abuse and dysfunctionalism: Current trends in phonology. *Linguistic Inquiry* 31, 157–169.
- Hale, Mark & Charles Reiss. 2008. *The Phonological Enterprise*. Oxford: Oxford University Press.
- Hauser, Marc D. 1996. *The Evolution of Communication*. Cambridge, MA: MIT Press.
- Hauser, Marc D., Susan Carey & Lillian B. Hauser. 2000. Spontaneous number representation in semi-free-ranging rhesus monkeys. *Proceedings of the Royal Society: Biological Sciences* 267, 829–833.
- Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002a. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569–1579.
- Hauser, Marc D., Elissa L. Newport & Richard N. Aslin. 2001. Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton top tamarins. *Cognition* 78, B53–B64.
- Hauser, Marc D., Daniel Weiss & Gary Marcus. 2002b. Rule learning by cotton-top tamarins. *Cognition* 86, B15–B22.
- Heinz, Jeffrey N. 2007. Inductive learning of phonotactic patterns. Los Angeles, CA: University of California dissertation.
- Hornstein, Norbert. 2001. *Move! A Minimalist Theory of Construal*. Malden, MA: Blackwell.
- Idsardi, William J. & Eric Raimy. In press. Three types of linearization & the temporal aspects of speech. In Teresa Biberauer & Ian Roberts (eds.), *Principles of Linearization*. Berlin: Mouton de Gruyter.
- Ishihara, Shinichiro. 2007. Major phrase, focus intonation, multiple spell-out (MaP, FI, MSO). *The Linguistic Review* 24, 137–167.
- Jackendoff, Ray & Steven Pinker. 2005. The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, & Chomsky). *Cognition* 97, 211–225.
- Johnson, C. Douglas. 1970. Formal aspects of phonological representation. Berkeley, CA: University of California dissertation.
- Johnson-Pynn, Julie, Dorothy M. Fragaszy, Elizabeth M. Hirsh, Karen E. Brakke & Patricia M. Greenfield. 1999. Strategies used to combine seriated cups by chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*), and capuchins (*Cebus apella*). *Journal of Comparative Psychology* 113, 137–148.
- Kahnemuyipour, Arsalan. 2004. The syntax of sentential stress. Toronto: University of Toronto dissertation.
- Kitahara, Hisatsugu. 2003. Acoustic perception in human newborns and cotton-top tamarins. *Reports of the Keio Institute of Cultural & Linguistic Studies* 35, 31–41.
- Kuhl, Patricia K. & James D. Miller. 1975. Speech perception by the chinchilla. *Science* 190, 69–72.
- Lakoff, George & Rafael Nuñez. 2001. *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*. New York: Basic Books.
- Leben, William. 1973. *Suprasegmental Phonology*. Cambridge, MA: MIT dissertation.

- Lepsius, Richard & William Dwight Whitney. 1865. On the relation of vowels & consonants. *Journal of the American Oriental Society* 8, 357–373.
- Levin, Juliette. 1985. *A Metrical Theory of Syllabicity*. Cambridge, MA: MIT dissertation.
- Mailhot, Frédéric & Charles Reiss. 2007. Computing long-distance dependencies in vowel harmony. *Biolinguistics* 1, 28–48.
- Marcus, Gary F., Sugumaran Vijayan, Shoba Bandi Rao, & Peter M. Vishton. 1999. Rule learning by seven-month-old infants. *Science* 283, 77–80.
- Merker, Björn. 2000. Synchronous chorusing and human origins. In Nils Wallin, Björn Merker & Steven Brown (eds.), *The Origin of Music*, 315–328. Cambridge, MA: MIT Press.
- Mielke, Jeff. 2008. *The Emergence of Distinctive Features*. Oxford: Oxford University Press.
- Mobbs, Iain. 2008. ‘Functionalism’, the design of the language faculty, and (disharmonic) typology. Ms., University of Cambridge.
- Feeleman, Ad & Hans van de Koot. 2006. On syntactic and phonological representations. *Lingua* 116, 1524–1552.
- Newport, Elissa L. & Richard N. Aslin. 2004. Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology* 48, 127–162.
- Newport, Elissa L., Marc D. Hauser, Geertrui Spaepen & Richard N. Aslin. 2004. Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology* 49, 85–117.
- Ohala, John J. 1992. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In *Papers from the CLS Parasession on the Syllable*, 319–338.
- Ohala, John J. & Haruko Kawasaki-Fukumori. 1997. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In Stig Eliasson & Ernst H. Jahr (eds.), *Language and Its Ecology: Essays in Memory of Einar Haugen*, 343–365. Berlin: Mouton de Gruyter.
- Ohshima, Nobuaki. 1997. Memorization of serial items by Japanese monkeys, a chimpanzee, and humans. *Japanese Psychological Research* 39, 236–252.
- Patel, Aniruddh. 2008. *Language and Music*. Oxford: Oxford University Press.
- Patel, Aniruddh & John R. Iversen. 2006. A non-human animal can drum a steady beat on a musical instrument. In *Proceedings of the 9th International Conference on Music Perception and Cognition*, 477.
- Payne, Katherine. 2000. The progressively changing songs of humpback whales: A window on the creative process in a wild animal. In Nils Wallin, Björn Merker & Steven Brown (eds.), *The Origins of Music*, 135–150. Cambridge, MA: MIT Press.
- Pinker, Steven & Ray Jackendoff. 2005. The faculty of language: What’s special about it? *Cognition* 95, 201–236.
- Raimy, Eric. 2000. *The Phonology and Morphology of Reduplication*. Berlin: Mouton de Gruyter.
- Raimy, Eric. 2003. Asymmetry and linearization in phonology. In Anna Maria Di Sciullo (ed.), *Asymmetry in Grammar*, vol. 2, 129–146. Amsterdam: John Benjamins.
- Ramus, Franck, Marc D. Hauser, Cory Miller, Dylan Morris & Jacques Mehler.

2000. Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science* 288, 349–351.
- Redford, Melissa A. 1999. An articulatory basis for the syllable. Austin, TX: University of Texas dissertation.
- Redford, Melissa A., Chun Chi Chen & Risto Miikkulainen. 2001. Constrained emergence of universals and variation in syllable systems. *Language and Speech* 44, 27–56.
- Reiss, Charles. 2008. The OCP and NOBANANA. 2008. In Bert Vaux & Andrew I. Nevins (eds.), *Rules, Constraints and Phonological Phenomena*, 252–301. Oxford: Oxford University Press.
- Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, MA: MIT Press.
- Saltzman, I.J. & W.R. Garner. 1948. Reaction time as a measure of span of attention. *Journal of Psychology* 25, 227–241.
- Samuels, Bridget. 2009a. The structure of phonological theory. Cambridge, MA: Harvard University dissertation.
- Samuels, Bridget. 2009b. Structure and specification in harmony. *Proceedings of NELS* 38(2), 283–296.
- Samuels, Bridget & Cedric Boeckx. 2009. What emerges from Merge in phonology? Paper presented at OCP 6, Edinburgh. [University of Edinburgh, 22–24 January 2009]
- de Saussure, Ferdinand. 1916. *Cours de linguistique générale*. Paris: Payot.
- Schusterman, Ronald J. & David Kastak. 1993. A California sea lion (*Zalophus californianus*) is capable of forming equivalence relations. *The Psychological Record* 43, 823–839.
- Seyfarth, Robert M., Dorothy L. Cheney & Thore J. Bergman. 2005. Primate social cognition and the origins of language. *Trends in Cognitive Sciences* 9, 264–266.
- Shettleworth, Sara J. 1998. *Cognition, Evolution, and Behavior*. Oxford: Oxford University Press.
- Slater, Peter J.B. 2000. Birdsong repertoires: Their origins and use. In Nils Wallin, Björn Merker & Steven Brown (eds.), *The Origins of Music*, 49–64. Cambridge, MA: MIT Press.
- Spelke, Elizabeth. 2003. What makes us smart? Core knowledge and natural language. In Dedre Gentner & Susan Goldin-Meadow (eds.), *Language in Mind: Advances in the Study of Language and Thought*, 277–311. Cambridge, MA: MIT Press.
- Smith, Dinitia. 1999. A thinking bird or just another birdbrain. *The New York Times*, October 9.
- Taylor, Ryan C., Barrett A. Klein, Joey Stein & Michael J. Ryan. 2008. Faux frogs: Multimodal signaling & the value of robotics in animal behavior. *Animal Behaviour* 76, 1089–1097.
- Toro, José M., Josep B. Trobalón & Núria Sebastián-Galles. 2005. The effects of backward speech & speaker variability in language discrimination by rats. *Journal of Experimental Psychology: Animal Behavior Processes* 31, 95–100.
- Truckenbrodt, Hubert. 1995. Phonological phrases: Their relation to syntax, focus, and prominence. Cambridge, MA: MIT Dissertation.
- Walter, Mary Ann. 2007. Repetition avoidance in human language. Cambridge,

- MA: MIT dissertation.
- Werker, Janet F. & Athena Voloumanos. 2000. Language: Who's got rhythm? *Science* 288, 280–281.
- Williams, Heather & Kirsten Staples. 1992. Syllable chunking in zebra finch (*Taeniopygia guttata*) song. *Journal of Comparative Psychology* 106, 278–286.
- Wright, Anthony A., Jacquelynne Rivera, Sara H. Hulse, Melissa Shyan & Julie Neiwirth. 2000. Music perception and octave generalization in rhesus monkeys. *Journal of Experimental Psychology: General* 129, 291–307.
- Yip, Moira. 2006a. Is there such a thing as animal phonology? In *Wondering at the Natural Fecundity of Things: Studies in Honor of Alan Prince*, online. [<http://repositories.cdlib.org/lrc/prince/15>.]
- Yip, Moira. 2006b. The search for phonology in other species. *Trends in Cognitive Sciences* 10, 442–445.

Bridget Samuels  
University of Maryland  
Department of Linguistics  
1401 Marie Mount Hall  
College Park, MD 20740  
USA  
[bridget@umd.edu](mailto:bridget@umd.edu)