What Complexity Differences Reveal About Domains in Language

Jeffrey Heinz,[a] William Idsardi[b]

[a]*Department of Linguistics and Cognitive Science, University of Delaware*
[b]*Department of Linguistics, University of Maryland*

Abstract

An important distinction between phonology and syntax has been overlooked. All phonological patterns belong to the regular region of the Chomsky Hierarchy, but not all syntactic patterns do. We argue that the hypothesis that humans employ distinct learning mechanisms for phonology and syntax currently offers the best explanation for this difference.

Keywords: Phonology; Syntax; Computational complexity; Language learning

1. A role for phonology in cognitive science

When it comes to the problem of how humans learn language, it appears that many computational learning theorists, cognitive scientists, and psychologists are primarily occupied with the problem of how humans learn to put words and morphemes together to form sentences. In this article, we argue that a further understanding of how sounds are put together to form words also bears directly on fundamental questions in cognitive science. In particular, we argue that computational analysis of the typology of patterns in phonology, when compared to the typology of patterns in syntax, reveals that cognitive learning mechanisms are likely multiple and modular in nature.

The skew that many researchers exhibit toward morpho-syntax may really be a skew toward studying meaning. But we believe that it is because phonological systems impose different sound patterns in different languages *without* contributing to meaning that they are especially interesting. That is, phonology is about "Rules without Meaning" in Frits Staal's (1989) terms.

Correspondence should be sent to Jeffrey Heinz, Department of Linguistics and Cognitive Science, University of Delaware, 125 E. Main Street, Newark, DE 19716. E-mail: heinz@udel.edu

We also believe that an apparent lack of teleological purpose in phonology is what lessens its appeal to the outside. A good discussion of the strangeness of studying phonology is provided in Kaye (1989), where he considers what a programming language like BASIC would look like if it had typical phonological processes such as assimilation. In this hypothetical language, the PRINT statement would assimilate to the first letter of the variable so that you would have statements such as PRIND D and PRINQ Q, but PRINT C would be ill-formed. Kaye asks if phonology is so important to human language, then why do not programming language designers include it? He goes on to suggest (as many others have) that the "purpose" of phonology is to smear the information in the signal across time; to add redundant bits like the extra error-correcting and detecting bits on compact discs. This Shannon-esque perspective is one useful way of looking at the problem, but we believe that it falls well short of the mark when long-distance patterns in phonology are considered. Long-distance patterns share important characteristics with the local assimilations, but as we discuss below, articulatory and/or perceptual-based explanations do not satisfactorily characterize their nature.

The results of the comparison between the computational complexity of phonology and syntax can be succinctly summarized. Phonological generalizations appear to require much less than context-free power, apparently staying within a small class of subregular patterns. Syntactic generalizations, on the other hand, appear to at least sometimes require something more than context-free power, appearing to converge on a mildly context-sensitive region of the Chomsky Hierarchy (see below).

What can explain this difference? We argue on the basis of long-distance dependencies in phonology that the explanation cannot be due to psychophysical constraints. On the other hand, if humans employ multiple, highly structured learning mechanisms in language learning, then the difference can be explained. In short, we put forward the hypothesis that humans generalize differently over words than they do over sentences.

We conclude with an analysis of the challenges faced by unitary language-learning algorithms, such as the one described in Chater and Vitányi (2007).

2. Phonology is different from syntax

Bromberger and Halle (1989) are often associated with the thesis that phonology differs from syntax. However, the arguments they put forward are quite different from the ones presented here. Our argument comes down to the formal character of the kinds of patterns languages display across sounds and words.

2.1. Pattern comparison

Patterns can be represented as sets of strings, relations, or as probability distributions over sets and relations. This fact provides a mathematically sound foundation for pattern analysis in any domain, be it bioinformatics (Durbin, Eddy, & Krogh Mitchison, 1998), robotic control and planning (Belta et al., 2007; Tanner et al., 2012), syntax (Chomsky, 1956, 1957), or phonology (Heinz, 2010a, 2011a,b).

In the realm of natural language, these sets and relations are infinite in size. However, grammars are precise, finite descriptions of these infinitely sized characterizations. Grammars allow one to enumerate the elements of the set or relation, and/or allow one to decide whether some string belongs to a set.

Theoretical computer science provides a way to analyze these mathematical objects. The Chomsky Hierarchy divides all logically possible patterns into nested regions of complexity (see Fig. 1). These regions have multiple, convergent definitions, which distill the necessary properties of any machine, grammar, or device that can recognize or generate the strings comprising the pattern (Harrison, 1978; Hopcroft, Motwani, & Ullman, 2001). It is worth emphasizing that these regions distinguish *abstract, structural* properties of grammars. For example, patterns belonging to the *regular* region are those for which devices only need *finitely* many internal states. Hence, it is said that regular patterns are those describable with finite-state grammars.

Importantly, there are probabilistic formulations for each of the regions in the Chomsky Hierarchy (Fig. 1). However, the choice of formulation matters (Kornai, 2011). To explain, stochastic languages are probability distributions over all logically possible strings. If regular languages are defined as those with regular support,[1] then, as Kornai explains, they may require full Turing Machine power to compute them so all distinctions are lost. On the other hand, if regular stochastic languages are defined to be those describable with finite-state machines augmented with probabilties on the transitions, then the traditional boundaries remain (Kornai, 2011, Theorem 2). Thus, the expressivity of stochastic and nonstochastic finite-state grammars is similarly limited because the key properties of "being regular" are independent of whether the function's codomain is real or boolean. It is this latter formulation we adopt when we speak of stochastic languages.

2.2. Syntax and phonology

A distinct advantage of this mathematical framework is its universality: it permits the comparison of patterns and their complexity across different domains. For example, in
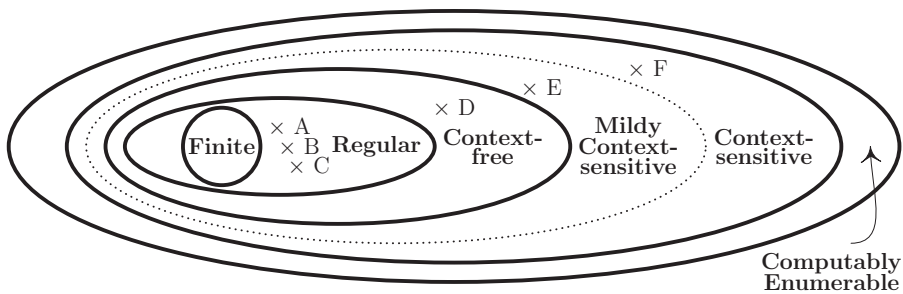

Fig. 1. The Chomksy Hierarchy. Point A represents English consonant cluster constraints (Clements & Keyser, 1983); B is sibilant harmony in Samala Chumash (Applegate, 1972); C is the unbounded stress pattern of Kwakiutl (Bach, 1975); D is English nested embedding (Chomsky, 1956); E is a dialect of Swiss German (Shieber, 1985); and F is copying constructions in Yoruba relative clauses (Kobele, 2006).

striking contrast to syntactic patterns, phonological patterns appear to belong to the *regular* region (Beesley & Kartunnen, 2003; Johnson, 1972; Kaplan & Kay, 1994). That syntax exhibits nonregular patterns was first argued by Chomsky (1956), and that syntax exhibits non-context-free patterns was argued by Shieber (1985). In other words, "being regular" is a universal property of phonological generalizations, but not of syntactic ones.

One obvious pattern that is nonregular is full reduplication. However, this is arguably a morphological process (Inkelas & Zoll, 2005; Roark & Sproat, 2007). While ultimately we must address the topic of morphology, the question of morphology is a large and controversial area, with many competing theories (Sadock, 1991; Anderson, 1992; Halle & Marantz, 1993), and therefore, we do not address it in this article. However, one approach to morphology consistent with our view would be to try to divide morpho-phonology from morpho-syntax. We would then expect morpho-phonology to be regular and aspects of morpho-syntax to be nonregular.

Because we attach great significance to the regular/nonregular distinction between phonology and syntax, it is useful to review the argument given by Kaplan and Kay (1994). (Johnson (1972) argues similarly, and Beesley & Kartunnen (2003) summarize the arguments in addition to providing background knowledge.) They show that optional, left-to-right, right-to-left, and simultaneous application of phonological rules (Chomsky and Halle, 1968) of the form A → B/C ___ D (where A,B,C,D are regular expressions) *describe regular relations*, provided the rule cannot reapply to the locus of its structural change.[2]

In Chomsky and Halle (1968), phonological generalizations are individual statements of the above type and phonological grammars are a list of such statements which apply so that the output of one rule is the input to the next. This is functional composition and regular relations are closed under functional composition. Therefore, any grammar which can be written in Chomsky and Halle's formalism describes a regular relation. As virtually all phonologies in the world's languages can be described with a list of rewrite rules, it follows that all phonologies in the world's languages are regular. Although SPE-style formalisms are no longer current, the fact that phonological generalizations can be expressed in this format means that the generalizations themselves are regular relations, even if they are expressed in different formalisms (e.g., Optimality Theory [Prince & Smolensky, 2004; see also Frank & Satta, 1998]). Similar considerations hold for phonotactic constraints, which are the right projections of the relations, which therefore are regular (and also closed under intersection).

To illustrate, consider the generalization that words in languages with sibilant harmony such as Navajo (Athabaskan) contain only anterior stridents (e.g., [s,z]) or nonanterior stridents (e.g., [ʃ,ʒ]) (Hansson, 2001; Sapir & Hojier, 1967). For example, words of the forms [… s… s…] are possible words of Navajo, but words of the forms [… s…ʃ…] and […ʃ… s …] are not.

This long-distance pattern can be described with a small finite-state acceptor as shown in Fig. 2. The acceptor encodes the long-distance process with states 1 and 2, which remember whether the word includes anterior (state 1) or nonanterior (state 2) stridents.

Returning to the broader point, the Chomsky Hierarchy has been used to classify patterns in a number of distinct domains, including Vedic ritual (Staal, 1989), robotics
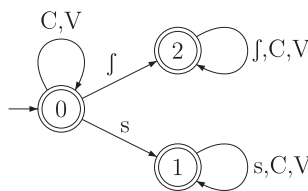
Fig. 2. A finite-state acceptor recognizing the sibilant harmony pattern of Navajo. Symbol [s] stands for any anterior strident; [ʃ] stands for any nonanterior strident; C stands for any nonstrident consonant; and V stands for any vowel.

(Rawal, Tanner, & Heinz, 2011), and biological sequences (Reinert, Schbath, & Waterman, 2005).

2.3. Subregular language classes

While "being regular" may be a *necessary* property of phonological generalizations, it is certainly not *sufficient*. There are many logically possible patterns lying in the regular region, which are not plausible phonological patterns. For example, consider a word-final stop deletion rule whose conditioning context is words with an even number of sibilant sounds. This is a logically possible, *regular* pattern. It is easy to think of many more such examples: phonological rules whose left contexts require the number of high vowels to equal 0 mod 3, or rules which require the first and last sibilants of a word to agree in the feature [anterior] (to the exclusion of word-medial sibilants).

The Subregular Hierarchies classify logically possible regular languages into nested regions of complexity (see Fig. 3). Like the regions in the Chomsky Hierarchy, these regions also have independently motivated converging characterizations from formal language theory, logic, abstract algebra, model theory, and automata theory (Heinz, Rawal, & Tanner, 2011; McNaughton & Papert, 1971; Rogers & Pullum, 2011; Rogers et al., 2010; Simon, 1975).
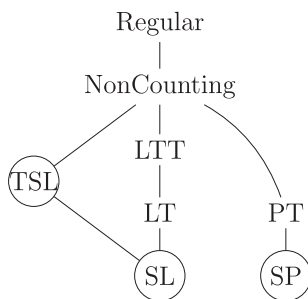

Fig. 3. Subregular classes of sets with proper inclusion relationships indicated from top to bottom. LT, locally testable; LTT, locally threshold testable; PT, piecewise testable; SL, strictly local; SP, strictly piecewise; TS, tier-based strictly local. Circled names indicate regions hypothesized to be where all individual phonotactic generalizations lie.

"Local" languages only make distinctions on the basis of *contiguous subsequences* up to some length k (called *k-factors*). A STRICTLY K-LOCAL (SL_k) grammar can be thought of as sets of prohibited k-factors and the language of the grammar are those words which do not contain any of those k-factors. For example, the grammar {*ab*} generates the language that contains all words with no contiguous *ab* sequence (i.e., *ab). Tier-based Strictly Local languages describe such constraints on phonological tiers (Heinz et al., 2011). Grammars of LOCALLY K-TESTABLE (LT_k) languages are finite boolean combinations of the SL_k languages. Thus, LT_2 languages can include words *bc* and *abcd*, but exclude *abc* (note this kind of distinction is not possible for SL_2 languages).

"Piecewise" languages only make distinctions on the basis of k-long *subsequences*. Subsequences are not necessarily contiguous; for example, *a…a* is a subsequence in the word *abcdab*. STRICTLY K-PIECEWISE (SP_k) and PIECEWISE K-TESTABLE (PT_k) languages can be defined analogously to SL_k and LT_k languages, respectively. For example, the SP grammar {*ab*} generates the language that contains all words with no subsequence *a…b* (i.e., **a…b*). Thus, SP and PT classes include patterns that describe long-distance dependencies, including those found in phonology like the Navajo example in Fig. 2 (Heinz, 2007, 2010a; Heinz & Rogers, 2010; Rogers et al., 2010).

Finally, the NONCOUNTING class contains those languages that do not count modulo some number n (so it excludes phonological constraints that penalize words with an even number of sibilants, for example). The regular languages that are not NonCounting, we call COUNTING.

Just as the regions in the Chomsky Hierarchy (Fig. 1) have probabilistic formulations, so do each of these subregular classes. Probabilistic SL_k languages are n-gram models (where $n = k$) (Garcia, Vidal, & Oncina, 1990), which are widely used in natural language processing (Jurafsky & Martin, 2008). In these models, the probability of the next symbol only depends on the previous contiguous sequence of length $n − 1$ (Markov, 1913). Heinz and Rogers (2010) define SP_k stochastic languages. These describe probability distributions over all strings such that the probability of the next symbol only depends on the set of previous (potentially discontiguous) subsequences of length $k − 1$.

Rogers and Pullum (2011) and Rogers et al. (unpublished data) argue that these subregular classes also have a cognitive status. For example, SL patterns require a short-term memory (see G1, Fig. 4). SP require a kind of long-term memory (see G2, Fig. 4). Finally, Counting patterns require a more general form of memory (see G3, Fig. 4).

It is important to recognize that the finite-state grammars in Fig. 4 do not obviously appear to distinguish any differences in the complexity of these different patterns, but the distinctions made by the Subregular Hierarchies do. According to the properties identified there, the SL and SP patterns are measurably *simpler* than Counting patterns. Furthermore, in stark contrast to the Counting patterns, the SL and SP classes include almost all phonotactic patterns (Heinz, 2010a).

The hypothesis that phonotactic constraints are either Strictly Local or Strictly Piecewise, with some limited exceptions, is supported by recent studies. For example, Hayes and Wilson (2008) adopt SL constraints, and many markedness constraints in Optimality Theory (Prince & Smolensky, 1993, 2004) appear to belong to SL (although no comprehensive
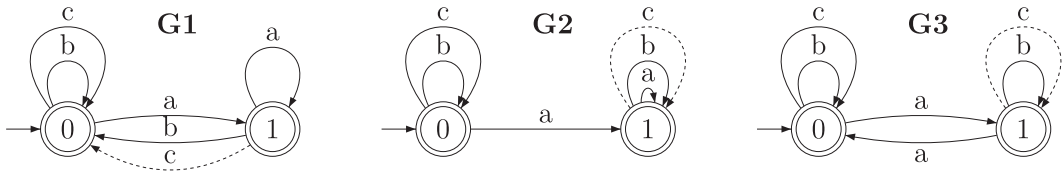
Fig. 4. Finite-state grammars for three regular patterns. G1 maintains a short-term memory w.r.t. [a] (i.e., State 1 means "just observed [a]"). G2 maintains a long-term memory w.r.t. [a] (i.e., State 1 means "observed [a] (sometime)"). G3 maintains a memory of the even/odd parity of [a]s (i.e., State 1 means "observed an even number of [a]s"). If dashed transitions are omitted, then G1 generates/recognizes all words except those with a forbidden string [ac]; G2 generates/recognizes all words except those with a forbidden subsequence [a… c]; and G3 generates/recognizes all words except those with a [c] whose left context contains an even number of [a]s. G1 is Strictly 2-Local, G2 is Strictly 2-Piecewise, and G3 is Counting.

survey has been conducted). Heinz (2010a) shows that long-distance phonotactic constraints derivable from long-distance consonant assimilation are Strictly Piecewise. As for stress patterns, Edlefsen et al. (2008) establish that 81 of 109 stress patterns studied by Heinz (2009) are Strictly Local. The other 28 patterns (which are unbounded stress patterns) may be Strictly Piecewise, once the generalization that all stress patterns have exactly one primary stress is factored out (Heinz, 2012). Another hypothesis is that phonotactic patterns are Tier-based Strictly Local; these were also used by Hayes and Wilson (2008). Very few languages, one being Creek, may exhibit Counting patterns in their stress systems (Graf, 2010) provided that the latent secondary stresses are in fact not perceptible.

2.4. Explaining the difference

We consider two hypotheses for explaining the complexity difference between phonology and syntax. One looks to external factors to constrain possible phonological patterns to be a subset of the possible syntactic patterns. The other directly distinguishes phonology from syntax by constraining phonological learning to subregular patterns.

The usual comment that we receive about why phonology is different is that it has to interface with input and output systems directly, whereas morphology and syntax do so only through phonology. At a general level we agree with this sentiment, but we find it difficult to operationalize this insight. Oversimplifying the intended proposals dramatically, they can be characterized in two broad ways: phonological generalizations are driven by motor planning and control issues (Browman & Goldstein, 1989; Gafos, 1999) or by general perceptual and memory constraints (Flemming, 1995; Ohala, 1981). The first we could call "coarticulation writ large," and the second "working memory effects" (in the sense that working memory is part of people's understanding of perception, especially when it shows expectation effects). From a learning perspective, these external factors could be thought to constrain phonological generalizations by interfering with a more powerful inference mechanism, which can also learn other language patterns such as syntactic ones.

We think that both kinds of external factors have real and obvious merits but fall short of providing a complete understanding of phonology. In particular, the nature of long-

distance patterns in phonology (and their parallels to local patterns) are puzzling from both of these points of view.

Consider first the "coarticulation writ large" position. Some analysts within phonology have proposed that long-distance effects are always mediated as local effects (Chiosáin & Padgett, 2001). For Navajo sibilant harmony, this would mean that all intervening segments between agreeing sibilants realize the feature [anterior]. Thus, such researchers would claim that an accurate transcription of [ʃiːteːʒ] "we (dual) are lying" would be [ʃiːteːʒ], where the underlined segments indicate a [-anterior] allophone of some kind, which is audible and perceptible at some level. With respect to the Subregular Hierarchies, this would have the effect of analyzing all SP patterns as SL ones.

However, Walker (1998) shows that this hypothesis does not hold up under careful phonetic analysis in all cases. Guaraní (Tupí) exhibits nasal harmony, which appears to spread across *oral* obstruents. For example, compare [ʃupa] "bed (1st poss.)" with [nũpã] "to hit." Walker's phonetic study confirms that the obstruents, which act transparent are indeed oral and *voiceless*. Walker's (1998) phonological analysis of this phenomenon is abstract, and in fact, any account of this generalization will have to similarly abstract away from the actual articulations of the Guaraní speakers. We conclude that long-distance generalizations cannot be analyzed as coarticulation run rampant.

If working memory effects were the explanation, then we should expect to see a fall-off in the force of these generalizations as the distance between the segments increases. There are reports of such fading generalizations both in experimental studies (Frisch, Pierrehumbert, & Broe, 2004) and in grammatical descriptions (Sapir & Hojier, 1967), but there are also cases where speakers' judgements are clear and strong across a wide range of distances. Perhaps the best evidence comes from examples of long words, which obey these generalizations in multiple languages (Hansson, 2001; Rose & Walker, 2004; Suzuki, 1998). For example, Samala (Chumash) has words like [ʃtojonowonowaʃ] "it stood upright" and [kʃalpujaťʃiʃi] "I get myself wet," but none like [ʃtojonowonowa**s**] nor [kʃalpujaťsiʃi] (Applegate,1972).

Furthermore, the general notions of working memory and perception are not concrete enough to account for both the patterns present in phonology and the patterns *absent* in phonology. For instance, it is well known that the first and last sounds of words are privileged positions and especially salient (Beckman, 1997; Endress, Nespor, & Mehler, 2009; Fougeron & Keating, 1997). These facts do not predict the patterns observed in Samala or Navajo. Instead, they favor constraints that require the first and last sounds of a word to agree in the feature [anterior] if they are sibilants (to the exclusion of word-medial sibilants). In this pattern, word-medial sibilants (being less salient) are not constrained at all. Words like [ʃtojonowonowaʃ, ʃtojono**s**onowaʃ, kʃalpujaťsiʃi] all obey this pattern. Despite the perceptual grounding of this pattern in privileged positions, it is striking to observe that no such phonological patterns or constraints have been reported; this pattern is completely unattested. Interestingly, this pattern, while regular, belongs to *neither* SL nor SP.

One potential source of evidence, which may help determine whether this pattern's absence from the typology is an accidental gap comes from artificial language-learning

experiments. Can humans generalize to the first–last harmony pattern in an artificial language setting? We discuss these experiments in the next section below.

While some future discovery may produce a theory of working memory and perception capable of distinguishing the subregular classes and grammars G1, G2, and G3, no such theory currently exists. We conclude that working memory and perception plays some role, but it is not itself sufficient to characterize the form of phonological generalizations. Moreover, appealing to abstract representations such as phonological tiers (Clements, 1985; Goldsmith, 1976; Mester, 1988) grants exactly the point we are trying to make: Phonological patterns are *abstractions* beyond a simple theory of coarticulation or working memory.

More generally, in the absence of some demonstration of the emergence of strictly piecewise effects from local effects of coarticulation and/or working memory, we believe that the difference noted in Fig. 1 and the taxonomy in Fig. 3 require a different explanation.

If humans employ multiple, highly structured learning mechanisms, then the difference in complexity between phonology and syntax can be explained. According to this hypothesis, humans generalize differently depending on whether they hear sounds in words versus words in sentences. When they hear words, two highly structured learning mechanisms are in play. One learns SL_k patterns and one learns $SP_{k'}$ patterns.

Formal studies of learnability reveal that both the SL_k and SP_k classes are feasibly learnable (Garcia et al., 1990; Heinz, 2010a, b) as are their probabilistic counterparts (Heinz & Rogers, 2010; Jurafsky & Martin, 2008). We emphasize that these learning mechanisms may not be specific to phonology but may apply in other domains as well, a point to which we return in Section 3.

Under this hypothesis, the form of phonological generalizations can be explained. Not only does what is attested follows but also what is unattested follows. The logically possible first–last harmony is unattested, for example, because even if humans were exposed to words conforming to this pattern, they would fail to detect it, as it lies outside the hypothesis spaces of humans' phonological "pattern detectors."

To summarize: phonology itself may have two or more different learning modules (Heinz, 2007, 2010a). While it is possible that syntax may have more than one learning mechanism as well, we do not address this possibility here. The next section explores the hypothesis that there are multiple learning modules for phonology in the context of established subregular language classes and artificial language learning experiments.

2.5. Artificial language-learning experiments

The basic assumption in artificial language-learning research is that some learning mechanisms are shared between artificial and natural language acquisition (Folia et al., 2010; Gómez & Gerken, 2000; Petersson, Forkstam, & Ingvar, 2004; Reber, 1967), so that the results obtained from these experiments inform natural language learning. Participants in these experiments are first exposed to stimuli (the artificial language) generated by some grammar. There are neither negative examples nor explicit feedback. Based on this exposure, can subjects internalize the rules defining the pattern present in the stimuli?

This is answered in the testing phase, wherein subjects are asked to respond to questions (which typically involve acceptability tasks), which reveal whether they have successfully abstracted the intended rule.

To date, few artificial language-learning experiments addressing phonology have focused on the formal character of the generalizations in the sense given by the discussion of the Subregular Hierarchies above. Instead, they have primarily focused on how strictly local generalizations are shaped by phonetic factors. Some research indicates that phonetic features do not appear to seriously constrain what is learnable in these settings (Chambers, Onishi, & Fisher, 2002; Cristiá & Seidl, 2008; Onishi, Chambers, & Fisher, 2003; Seidl & Buckley, 2005). On the other hand, some do find biases in vowel harmony patterns (Finley & Badecker, 2009a, b; Moreton, 2008; Pycha, Nowak, Shin, & Shosted, 2003). Wilson (2006) presents experimental evidence, which also supports the hypothesis that humans are substantively biased toward patterns that are phonetically natural, a viewpoint compatible with Bayesian approaches (Griffiths, Kemp, & Tenenbaum, 2008).

Finley (2011, 2012) examines long-distance consonantal harmony patterns in artificial language-learning experiments. Her findings complement the typological evidence regarding the long-distance nature of these patterns. Experiments by Koo and Callahan (2012) test the hypothesis that long-distance patterns are necessarily phonetically natural and find that long-distance, phonetically unnatural patterns can be learned. The results of these experiments are consistent with the SP hypothesis space.

However, only one experiment to date has examined whether logically possible phonological patterns outside the SL and SP regions can be learned in artificial language-learning experiments. Lai (2012) uses artificial language-learning experiments to compare the learnability of sibilant harmony patterns like the ones found in Navajo and Samala, with the unattested first–last pattern. If subjects learn both patterns equally well, this would constitute evidence against the specific claim that the phonological learning mechanism could not learn patterns outside the SL or SP regions.

Artificial language-learning experiments can also test directly whether the computational difference between syntax and phonology is relevant to learning (Lai, 2012). Fig. 5 shows how nonregular patterns attested in syntax can be embedded into words.

Do subjects internalize these patterns as indicated by their behavior on novel forms during testing? As before, the hypothesis that the such nonregular patterns are learnable
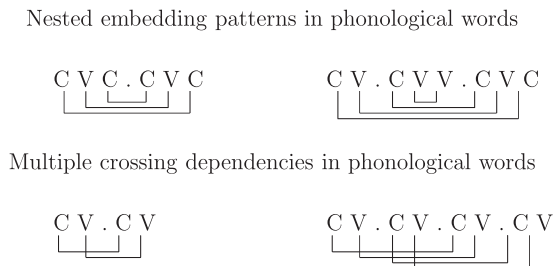

Fig. 5. Nonregular patterns in words. Lines indicate agreement along particular phonetic features, such as voicing or vowel height.

only if they are presented in sentences would predict that subjects could internalize such patterns if they were presented as words and morphemes within sentences, but not if they are presented as sounds within words (as shown in Fig. 5).

There is precedent for this kind of study. Bergelson and Idsardi (2009) examine whether two phonological patterns of the same formal character are equally learnable when one is instantiated over segments and the other is instantiated over syllables as a stress pattern. They report that how the pattern is instantiated does matter.

In sum, artificial language-learning experiments can probe the boundaries of proposed hypothesis spaces. By comparing the learnability of patterns on either side of some proposed boundary, evidence can be collected in favor or against specific proposals of how humans learn language. These experiments, however, do not directly address whether the learning models are domain specific or domain general (but again see Lai (2012)).

3. Domain-general and domain-specific algorithms

Domain-specific and domain-general ought not be confused with small (highly structured) concept classes and large (unstructured) concept classes, respectively. Rather they mean "applies in only one domain" and "applies in more than one domain," respectively. Thus, whether a model is domain general or domain specific is not due to any inherent property of the model itself.

Furthermore, there are two senses in which a model might be said to be domain general or domain specific. On the one hand, a model may be considered to be domain general because in truth the model applies to more than one distinct domain. On the other hand, a model may be considered to be domain general because researchers apply it to several domains, regardless of what the truth is. Similar statements can be made with respect to domain-specific models.

The domain generality or specificity of a model ought to be determined empirically in both senses of the terms. It is easy to decide if a model is domain specific or domain general by seeing to which domains researchers apply such models. While it is more difficult to decide if a model for a domain is "the truth," it ought to be clear that if a model is "the truth" for more than one domain, then it is clearly a domain-general model.

These distinctions are illuminated in Perfors, Tenenbaum, and Regier (2011). On page 308, they provide a useful figure, which is reproduced here as Fig. 6.

Perfors et al. (2011) suggest that until the present, most models fall into either areas labeled A and D, and that domain-general, structured models (area B) have been left largely unstudied.

However, there is a highly structured, domain-general model in wide use which they have overlooked. The n-gram model is probably the best example of a richly structured, domain-general model. It is richly structured because its hypothesis space is tightly constrained. As mentioned above, they are the statistical counterpart to the SL class. Most regular patterns are not strictly local. It is domain-general in both senses described above. Researchers apply n-gram models in a range of tasks within natural language processing

Structured

A   B

Domain-specific ◄━━━━━━━━━━━━━━► Domain-general
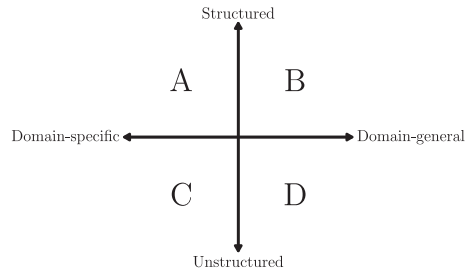
C   D

Unstructured

Fig. 6. Figure showing space of learning algorithms along two dimensions. Adapted from Fig. 1 from Perfors et al. (2011, p. 308).

(Jurafsky & Martin, 2008) as well as to tasks in visual processing (Turk-Browne, Scholl, Chun, & Johnson, 2009). N-gram models also plausibly play a role in many domains where the next observable depends on the immediately preceding observable.

Perfors et al. (2011) claim their model is domain-general. It certainly invokes principles that seem applicable to more than one domain. It also invokes elements that appear specific to language; for example, as described in the appendix, it comes pre-equipped with syntactic categories. So it is unclear to us whether their specific proposal would fall on the left or right side of the diagram. Clearly, this can be settled in their favor by demonstrating the model's success in some nonlinguistic domain.

With respect to the unstructured, domain-general area in Fig. 6, we place algorithms capable of learning any stochastic context-free or computably enumerable distribution (Angluin, 1988; Chater & Vitányi, 2007; Horning, 1969) in this area.

4. How ideal is "ideal" language learning?

Chater and Vitányi (2007) prove that in a particular learning setting, any computably enumerable distribution can be learned from positive evidence. Their result goes beyond similar results by Angluin (1988) and Horning (1969) in that the stream of linguistic data to which the learner is exposed need not come from a fixed-probability distribution, but from a nonstationary one (although it must still be generable by Turing machine). A core idea in this work is that "the learner postulates the underlying structure in the linguistic input that provides the simplest, that is, briefest, description of that linguistic input." In other words, their learner makes use of the notion of Minimum Description Length (Grünwald, 2007).

Chater and Vitányi (2007) offer some caveats about their ideal language learner in both their introduction and conclusion. For example, page 308 explains that their learner makes "calculations that are known to be uncomputable" and that they do not "directly address the question of the speed of learning." These particular concerns, while important, are not the focus here. In fact, the arguments here would be made even if the ideal language learner was computable in log-linear time.

Instead, we consider their proposal as one of how humans learn language[3] and address the following questions.

1. In models of human language acquisition, is it desirable for the model to be able to learn any computably enumerable pattern?
2. In models of human language acquisition, is it even desirable for the model to be able learn any finite language?
3. In models of human language acquisition, is it desirable to make hypotheses on the basis of Minimum Description Length?

In this section, we argue that the correct answer to these questions is No.

One reason to think that not any comptably enumerable language is learnable is because the kinds of linguistic generalizations attested in the world's languages are not arbitrary. The nonarbitrary nature of linguistic generalizations strongly indicates the presence of universal principles, which limit the space of logically possible generalizations humans make when encountering linguistic data. Indeed, language typologists repeatedly observe that the extensive variation that exists appears to be limited, although stating exact universals is difficult (Greenberg, 1963, 1978; Mairal & Gil, 2006; Stabler, 2009). While some conclude that there are, therefore, no universals (Evans & Levinson, 2009), others have argued that the nature of the universals is necessarily abstract (Keenan & Stabler, 2004, 2009). And even Evans and Levinson (2009) acknowledge that natural language patterns can only ever occupy a small corner of the "design space" (p. 478), although they would rather attribute this to functional considerations than grammatical ones.

We suspect that not even any finite language is learnable by humans. One argument comes from a thought experiment that Tenenbaum presents regularly to his audiences and which is described on page 1,280 of Tenenbaum, Kemp, Griffiths, and Goodman (2011) (see Fig. 7). Tenenbaum observes that when people are informed that the items in red boxes are "tufas" that they infer that the items in gray boxes are also "tufas," even though none of the items are identical. Suppose that in fact the only "tufas" were exactly the three items in the red boxes! We suspect Tenenbaum has never encountered anyone in his audiences who thought the only "tufas" were the objects in the red boxes. This possibility is so outside of what humans are willing to conceive that it would be somewhat shocking (and disappointing) to realize that the class of "tufas" is exactly the three items observed. One would feel tricked.

One objection to this thought experiment is that asking people which other items are "tufas" presupposes that other items are "tufas." But, simple modifications to the experiment eliminate this objection. For example, suppose for your next three meals you eat each of the items in the red boxes and they each taste bad. At subsequent mealtimes, you will probably be disappointed to receive items in the gray boxes because you will suspect they will be bad tasting in contrast to the other items, which you are much more likely to approach with an open mind. The impetus for humans to generalize beyond their finite experience is very strong.

Proponents of MDL to human language learning could take these observations as evidence in favor of MDL. After all, they could claim that representing three items is more

Fig. 7. Figure 1a from page 1,280 of Tenenbaum et al. (2011).

complex than the grammatical description that represents some class of items. As MDL is a mathematically precise theory, it is possible to identify exactly when a grammatical description is shorter than a list of individual elements. So when is it? And Chater and Vitányi (2007) have proved that any finite class of languages is learnable by their ideal language learner. How much and what kind of input data is needed to learn each finite language?

The observations above are not so much evidence in favor of MDL as they are evidence in favor of humans not maintaining finite-list hypotheses about their experience. This is illustrated by the many non-MDL learning algorithms, which generalize immediately on the basis of a few examples. Unsmoothed n-gram models provide a simple example. A bigram model that observes the single word *aa* immediately assigns nonzero probabilities to words a^n, $n > 1$. In fact, the SL_k and SP_k regions are learnable in part because they exclude many finite languages.

Lastly, we are skeptical that MDL is the right basis for generalization, at least in phonology. Recall the subregular classes shown in Fig. 3. If MDL were the right basis for generalization, then we might expect that patterns which belong to the circled classes have shorter description lengths because these are the classes which describe most phonological patterns. However, in Fig. 4 no such correlation holds. The FSAs in Fig. 4 are the smallest descriptions of these patterns, and one is Strictly Local, one is Strictly Piecewise, and one is Counting. In fact, if we consider another Strictly Local pattern like G1, but

which instead bans *aac* sequences in words, then this pattern would have a minimal description length which includes three states, whereas the Counting pattern given by G3 has only two states (and would have fewer transitions).

This argument is not a deductive proof that there is no formal system that corresponds to the subregular complexity hierarchy. It is well known that there are multiple ways regular languages can be represented such as with regular expressions (Hopcroft, Motwani, & Ullman, 2001), as solutions to certain algebraic equations (Kracht, 2003), with their syntactic monoids (McNaughton & Papert, 1971) or universal automata (Lombardy & Sakarovitch, 2008), among many more. Importantly, relative length of generalization is not preserved across these formalisms and so it is unknown which of these if any correspond to the complexity hierarchies.[4]

In other words, as far as we know, the subregular hierarchies describe computational complexity in different terms than MDL. While MDL is flexible enough to choose some other universal representational scheme, we know of no such scheme that correlates with the arrangement of classes shown in Fig. 3. Progress in MDL formulations will require investigations of these formal systems and comparing them to the subregular hierarchies.

Whether the subregular complexity hierarchies are actually appropriate can be settled empirically by designing a series of artificial language-learning experiments which, for example, compare the learnability of a SL_3 patterns with patterns which count mod 2. More sophisticated learning paradigms such as iterated learning (Kalish et al., 2007; Scott-Phillips & Kirby, 2010) may also shed some light on this question.

The ideal language learner is just one example of a highly unstructured learning algorithm. It is unstructured because there are no limits on its hypothesis space apart from that the patterns are describable with a Turing machine. We agree with Perfors et al. (2011) that structured learning models are more promising than unstructured models for language learning. It is telling that Tenenbaum et al. (2011) also recognize problems, which arise with unstructured learning models. In their efforts to scale up the Bayesian models they currently employ, Tenenbaum et al. (2011, p. 1284) write

> The most daunting challenge is that formalizing the full content of intuitive theories appears to require Turing-complete compositional representations, such as probabilistic first-order logic and probabilistic programming languages. How to effectively constrain learning with such flexible representations is not at all clear.

In addition to the arguments we have made here, the above questions can be turned into empirical ones as follows.

1. Setting aside performance considerations, can humans learn any computably enumerable pattern?
2. Setting aside performance considerations, can humans learn any finite language?
3. When learning language, do humans generalize on the basis of Minimum Description Length?

As discussed earlier, the artificial language-learning experimental paradigm can shed some light on these empirical questions.

5. Discussion and conclusion

Nick Chater helpfully distinguishes three possible factors in building an explanation for the observed difference between phonology and syntax (this is not meant to be exhaustive).

1. Phonology and syntax have different roles and the observed difference are shaped by functional factors relating to those roles.
2. Phonology and syntax draw differentially on different cognitive machinery.
3. Phonology and syntax draw on different learning mechanisms.

These three possibilities could all be true, simultaneously. However, we fail to see the linking hypotheses between 1 and 2 to the regular/nonregular distinction. On the other hand, the position we have adopted (3) makes explicit correspondences between attested and unattested linguistic generalizations and learners that can and cannot learn them.

The comparison between syntax and phonology reveals a major computational difference in the sorts of generalizations in those domains. The difference between these two domains is well characterized by the Chomsky and Subregular Hierarchies. This kind of analysis is applicable beyond language and, therefore, in this sense the analysis we have made is domain general. On the other hand, we argued for multiple, distinct learning mechanisms within language, and by extension, for domains distinct from language. We expect some problems will show subregular patterns while others will show full recursive structures. In the domain of songbird learning, for example, most people have taken birdsong to be a syntax-like phenomenon, but Berwick et al., (2011) analyze the songs as being more similar in character to human phonology (i.e., subregular).

Accounting for the difference between phonology and syntax must not only explain what occurs but also what does not occur. As the old saying goes, "People are happy not only because they have what they want but also because they don't have what they don't want" (such as cancer). The hypothesis that there are highly structured, restricted learning mechanisms for phonology is currently the best explanation for both what phonology contains, for what it does not contain, and hence for why it is different from syntax.

Acknowledgments

Notes

1. This means the language given by the strings with nonzero probabilities is regular.
2. This means rules can reapply to their own output, provided the part of the string currently targeted by the rule is not properly contained in what was just rewritten.
3. Chater's and Vitányi's (2007) primary goal in their article was to show that language learning is possible from positive data only.
4. Rogers et al. (to appear) establish that regular expressions also fail to make the distinctions present in the subregular hierarchies.

References

Anderson, S. R. (1992). *A-morphous morphology*. Cambridge, England: Cambridge University Press.

Angluin, D. (1988). *Identifying languages from stochastic examples*. Tech. Rep. 614, New Haven, CT: Yale University.

Applegate, R. B. (1972). *Ineseño Chumash grammar*. Doctoral dissertation, Berkeley: University of California.

Bach, E. (1975). Long vowels and stress in Kwakiutl. *Texas Linguistic Forum*, 2, 9–19.

Beckman, J. (1997). Positional faithfulness, positional neutralization, and Shona vowel harmony. *Phonology*, 14, 1–46.

Beesley, K., & Kartunnen, L. (2003). *Finite state morphology*. Stanford, CA: CSLI Publications.

Belta, C., Bicchi, A., Egerstedt, M., Frazzoli, E., Klavins, E., & Pappas, G. J. (2007). Symbolic planning and control of robot motion. *IEEE Robotics and Automation Magazine*, 14, 61–71. Special issue on grand challenges for robotics.

Bergelson, E., & Idsardi, W. J. (2009). Structural biases in phonology: Infant and adult evidence from artificial language learning. In J. Chandlee, M. Franchini, S. Lord, & G.-M. Rheiner (Eds.), *BUCLD 33: Proceedings of the 33rd annual Boston university conference on language development* (pp. 85–96). Somerville, MA: Cascadilla Press.

Berwick, R. C., Okanoya, K., Beckers, G. J. L., & Bolhuis, J. J. (2011). Songs to syntax: The linguistics of birdsong. *Trends in Cognitive Sciences*, 15, 113–121.

Bromberger, S., & Halle, M. (1989). Why phonology is different. *Linguistic Inquiry*, 20, 51–70.

Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6, 201–251.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, 83, B13–B23.

Chater, N., & Vitányi, P. (2007). "Ideal learning" of natural language: Positive results about learning from positive evidence. *Journal of Mathematical Psychology*, 51, 135–163.

Chiosáin, M. N., & Padgett, J. (2001). Markedness, segment realization, and locality in spreading. In L. Lombardi (Ed.), *Constraints and representations: Segmental phonology in optimality theory* (pp. 118–156). Cambridge, England: Cambridge University Press.

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory*, 2(3), 113–124. IT-2.

Chomsky, N. (1957). *Syntactic structures*. The Hague, The Netherlands: Mouton & Co., Printers.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Clements, G. N., & Keyser, S. J. (1983). *CV phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.

Clements, G. N. (1985). The geometry of phonological features. *Phonology Yearbook*, 2, 225–252.

Cristiá, A., & Seidl, A. (2008). Phonological features in infants phonotactic learning: Evidence from artificial grammar learning. *Language, Learning, and Development*, *4*, 203–227.

Durbin, R., Eddy, S. R., Krogh, A., & Mitchison, G. (1998). *Biological sequence analysis: Probabilistic models of proteins and nucleic acids*. Cambridge, UK: Cambridge University Press.

Edlefsen, M., Leeman, D., Myers, N., Smith, N., Visscher, M., & Wellcome, D. (2009). Deciding strictly local (SL) languages. In J. Breitenbucher (Ed.), *Proceedings of the midstates conference for undergraduate research in computer science and mathematics* (pp. 66–73). Oberlin, OH: Oberlin College.

Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Science*, *13*, 348–353.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*, 429–492.

Finley, S. (2011). The privileged status of locality in consonant harmony. *Journal of Memory and Language*, *65*, 74–83.

Finley, S. (2012). Testing the limits of long-distance learning: Learning beyond the three-segment window. *Cognitive Science*, *36*, 740–756.

Finley, S., & Badecker, W. (2009a). Artificial language learning and feature-based generalization. *Journal of Memory and Language*, *61*, 423–437.

Finley, S., & Badecker, W. (2009b). Right-to-left biases for vowel harmony: Evidence from artificial grammar. In A. Shardl, M. Walkow, & M. Abdurrahman (Eds.), *Proceedings of the 38th North East linguistic society annual meeting*, vol. *1* (pp. 269–282). Amherst, MA: University of MA, Amherst GLSA.

Flemming, E. S. (1995). *Auditory representations in phonology*. Doctoral dissertation, Los Angeles: UCLA.

Folia, V., Uddén, J., de Vries, M., Forkstam, C., & Petersson, K. M. (2010). Artificial language learning in adults and children. *Language Learning*, *60*(Supplement 2), 188–220.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustic Society of America*, *101*, 3728–3740.

Frank, R., & Satta, G. (1998). Optimality theory and the generative complexity of constraint violability. *Computational Linguistics*, *24*, 307–315.

Frisch, S., Pierrehumbert, J., & Broe, M. (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory*, *22*, 179–228.

Gafos, A. (1999). *The articulatory basis of locality in phonology*. New York: Garland.

Garcia, P., Vidal, E., & Oncina, J. (1990). Learning locally testable languages in the strict sense. In *Proceedings of the Workshop on Algorithmic Learning Theory*, 325–338.

Goldsmith, J. (1976). *Autosegmental phonology*. Doctoral dissertation, Cambridge, MA: MIT. Published by Garland Press, New York, 1979.

Gómez, R. L., & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*, 178–186.

Graf, T. (2010). Comparing incomparable frameworks: A model theoretic approach to phonology. University of Pennsylvania Working Papers in Linguistics 16: Article 10. Available at: http://repository.upenn.edu/pwpl/vol16/iss1/10. Accessed December 5, 2012.

Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of Language*, 73–113. Cambridge, MA: MIT Press.

Greenberg, J., ed. (1978). *Universals of human language: Volume 2, phonology*. Stanford, CA: Stanford University Press.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge handbook of computational cognitive modeling* (pp. 59–100). New York: Cambridge University Press.

Grünwald, P. (2007). *The minimum description length principle*. Cambridge, MA: MIT Press.

Halle, M., & Marantz, A. (1993). Distributed morphology. In K. Hale & S. J. Keyser (Eds.), *The view from building 20. Essays in honor of Sylvain Bromberger* (pp. 111–176). Cambridge, Mass.: MIT Press.

Hansson, G. (2001). *Theoretical and typological issues in consonant harmony*. Doctoral dissertation, Berkeley: University of California.

Harrison, M. A. (1978). *Introduction to formal language theory*. Boston, MA: Addison-Wesley Publishing Company.

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, *39*, 379–440.

Heinz, J. (2007). *The inductive learning of phonotactic patterns*. Doctoral dissertation, Los Angeles: University of California.

Heinz, J. (2009). On the role of locality in learning stress patterns. *Phonology*, *26*, 303–351.

Heinz, J. (2010a). Learning long-distance phonotactics. *Linguistic Inquiry*, *41*, 623–661.

Heinz, J. (2010b). String extension learning. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 897–906). Uppsala, Sweden: Association for Computational Linguistics.

Heinz, J. (2011a). Computational phonology part I: Foundations. *Language and Linguistics Compass*, *5*, 140–152.

Heinz, J. (2011b). Computational phonology part II: Grammars, learning, & the future. *Language and Linguistics Compass*, *5*, 153–168.

Heinz, J. (2012). *Culminativity times harmony equals unbounded stress*. Cambridge, UK: Cambridge University Press. MS. To appear in volume edited by Harry van der Hulst from the April 2010 StressTyp Workshop.

Heinz, J., Rawal, C., & Tanner, H. (2011). Tier-based strictly local constraints for phonology. In *Proceedings of the 49th annual meeting of the association for computational linguistics* (pp. 58–64). Portland, Oregon, USA: Association for Computational Linguistics.

Heinz, J., & Rogers, J. (2010). Estimating strictly piecewise distributions. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 886–896). Uppsala, Sweden: Association for Computational Linguistics.

Hopcroft, J., Motwani, R., & Ullman, J. (2001). *Introduction to automata theory, languages, & computation*. Boston, MA: Addison-Wesley.

Horning, J. J. (1969). *A study of grammatical inference*. Doctoral dissertation, Stanford, CA: Stanford University.

Inkelas, S., & Zoll, C. (2005). *Reduplication: Doubling in morphology*. New York, NY: Cambridge University Press.

Johnson, C. D. (1972). *Formal aspects of phonological description*. The Hague, The Netherlands: Mouton.

Jurafsky, D., & Martin, J. (2008). *Speech and language processing: An introduction to natural language processing, speech recognition, and computational linguistics*. 2nd ed. Upper Saddle River, NJ: Prentice-Hall.

Kaplan, R., & Kay, M. (1994). Regular models of phonological rule systems. *Computational Linguistics*, *20*, 331–378.

Kaye, J. (1989). *Phonology: A cognitive view*. Hillsdale, NJ: L. Erlbaum Associates.

Keenan, E., & Stabler, E. (2004). *Bare grammar*. Stanford, CA: Stanford University Press.

Keenan, E., & Stabler, E. (2009). Language variation and linguistic invariants. *Lingua*, *120*, 2680–2685.

Kobele, G. (2006). *Generating copies: An investigation into structural identity in language and grammar*. Doctoral dissertation, Los Angeles: University of California.

Koo, H., & Callahan, L. (2012). Tier-adjacency is not a necessary condition for learning phonotactic dependencies. *Language and Cognitive Processes*, *27*(10), 1425–1432.

Kornai, A. (2011). Probabilistic grammars and languages. *Journal of Logic, Language, and Information*, *20*, 317–328.

Kracht, M. (2003). *The mathematics of language*. Berlin: Mouton de Gruyter.

Lai, R. (2012). Domain Specificity in Learning Phonology. Doctoral dissertation, Newark, DE: University of Delaware.

Lombardy, S., & Sakarovitch, J. (2008). The universal automaton. In J. Flum, E. Grädel, & T. Wilke (Eds.), *Logic and automata,* vol. 2 of *Texts in Logic and Games* (pp. 457–504). Amsterdam: Amsterdam University Press.

Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin and Review*, *14*, 288–294.

Mairal, R., & Gil, J. (eds.) (2006). *Linguistic universals*. New York: Cambridge University Press.

Markov, A. A. (1913). An example of a statistical study on the text of "Eugene Onegin", illustrating connecting observations into a chain. Annals of the Russian Imperial Academy of Science. *x*, no. 3, 1913.

McNaughton, R., & Papert, S. (1971). *Counter-free automata*. Cambridge, MA, MIT Press.

Mester, A. (1988). *Studies in Tier structure*. New York: Garland Publishing, Inc.

Moreton, E. (2008). Analytic bias and phonological typology. *Phonology*, *25*, 83–127.

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrik, & M. F. Miller (Eds.), *Papers from the parasession on language and behavior* (pp. 178–203). Chicago: Chicago Ling. Soc.

Onishi, K. H., Chambers, K. E., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, *87*, B69–B77.

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, *118*, 306–338.

Petersson, K. M., Forkstam, C., & Ingvar, M. (2004). Artificial syntactic violations activate brocas region. *Cognitive Science*, *28*, 383–407.

Prince, A., & Smolensky, P. (1993). Optimality theory: Constraint interaction in generative grammar. Tech. Rep. 2, *Rutgers University Center for Cognitive Science*, New Brunswick, NJ: Rutgers University.

Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell Publishing.

Pycha, A., Nowak, P., Shin, E., & Shosted, R. (2003). Phonological rule-learning and its implications for a theory of vowel harmony. In G. Garding & M. Tsujimura (Eds.), *Proceedings of the West Coast Conference on Formal Linguistics 22* (pp. 423–435). Somerville, MA: Cascadilla Press.

Rawal, C., Tanner, H. G., & Heinz, J. (2011). (Sub)regular robotic languages. In *IEEE Mediterranean Conference on Control and Automation*, 321–326.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*, 855–863.

Reinert, G., Schbath, S., & Waterman, M. (2005). Statistics on words with applications to biological sequences. In M. Lothaire (Ed.), *Applied combinatorics on words*. Cambridge, UK: Cambridge University Press.

Roark, B., & Sproat, R. (2007). *Computational approaches to morphology and syntax*. Oxford: Oxford University Press.

Rogers, J., Heinz, J., Bailey, G., Edlefsen, M., Visscher, M., Wellcome, D., & Wibel, S. (2010). On languages piecewise testable in the strict sense. In C. Ebert, G. Jäger, & J. Michaelis (Eds.), *The Mathematics of Language,* vol. 6149 of *Lecture Notes in Artifical Intelligence* (pp. 255–265). Berlin: Springer.

Rogers, J., Heinz, J., Fero, M., Hurst, J., Lambert, D., & Wibel, S. (to appear). Cognitive and sub-regular complexity. In *Proceedings of the 17th Conference on Formal Grammar*.

Rogers, J., & Pullum, G. (2011). Aural pattern recognition experiments and the subregular hierarchy. *Journal of Logic, Language and Information*, *20*, 329–342.

Rose, S., & Walker, R. (2004). A typology of consonant agreement as correspondence. *Language*, *80*, 475–531.

Sadock, J. (1991). *Autolexical syntax*. Chicago: University of Chicago Press.

Sapir, E., & Hojier, H. (1967). *The phonology and morphology of the Navaho language*. Berkeley, CA: University of California Publications 50.

Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences*, *14*, 411–417.

Seidl, A., & Buckley, E. (2005). On the learning of arbitrary phonological rules. *Language Learning and Development*, *1*, 289–316.

Shieber, S. (1985). Evidence against the context-freeness of natural language. *Linguistics and Philosophy*, 8, 333–343.

Simon, I. (1975). Piecewise testable events. In *Automata Theory and Formal Languages*, H. Barkhage (Ed.). Lecture Notes in Computer Science 33 Springer, 214–222.

Staal, F. (1989). *Rules without meaning: Ritual, mantras and the human sciences*. New York: Peter Lang.

Stabler, E. P. (2009). Computational models of language universals. In C. Collins & S. Edelman, *Language universals* (pp. 200–223). Oxford, England: Oxford University Press.

Suzuki, K. (1998). *A typological investigation of dissimilation*. Doctoral dissertation, Tucson, AZ: University of Arizona.

Tanner, H. G., Rawal, C., Fu, J., Piovesan, J. L., & Abdallah, C. T. (2012). Finite abstractions for hybrid systems with stable continuous dynamics. *Discrete Event Dynamic Systems*, 22, 83–99.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure and abstraction. *Science*, 331, 1279–1285.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, 21, 1934–1945.

Walker, R. (1998). *Nasalization, neutral segments, & opacity effects*. Doctoral dissertation, Santa Cruz: University of California.

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, 30, 945–982.