2012

Structure and Substance in Artificial-Phonology Learning, Part I: Structure

Elliott Moreton, *University of North Carolina at Chapel Hill*
Joe Pater

Structure and substance in artificial-phonology learning, Part I: Structure

Elliott Moreton — UNC-Chapel Hill
Joe Pater — UMass-Amherst *

Version 5.0, June 26, 2012

Abstract

Artificial analogues of natural-language phonological patterns can often be learned in the lab from small amounts of training or exposure. The difficulty of a featurally-defined pattern has been hypothesized to be affected by two main factors, its formal structure (the abstract logical configuration of the defining features) and its phonetic substance (the concrete phonetic interpretation of the pattern). This paper, the first of a two-part series, reviews the experimental literature on structural effects. The principal finding is a robust complexity effect: Patterns which depend on more features are reliably harder to learn.

Contents

*The authors gratefully acknowledge the advice and suggestions already contributed by two anonymous reviewers, Bruce Hayes, Anne Pycha, Jen Smith, Kie Zuraw, and the participants in Linguistics 751 at the University of Massachusetts in Spring 2011, especially John Kingston, Claire Moore-Cantwell, Wendell Kimper, and Robert Staubs. This paper is Part I of a two-paper series, and is intended to be read in conjunction with Part II.

1 Introduction

Laboratory study of the acquisition of artificial phonological patterns is potentially very significant as a window on the inductive biases involved in acquiring natural-language phonology.[1] Some such biases must exist, since generalizations cannot be learned without them (e.g., Pinker, 1979; Mitchell, 1990; Gallistel et al., 1991); what is at issue is rather their content and their causal relationship to phenomena of natural language acquisition, change, and typology. Artificial phonology may offer a way to study these biases in near-isolation from each other and from other confounding factors. The present article (Part I) and its companion (Part II) review the empirical literature on artificial-phonology learning in the context of this program.

What factors make phonological patterns harder to learn? And do these highly artificial tasks reveal anything about natural-language phonology? Two hypothesized biasing factors have been studied the most intensively, *formal complexity* and *phonetic substance*. A learner with formal complexity bias would acquire simpler patterns faster or better than complex ones, whereas a learner with substantive bias would acquire phonetically-motivated patterns better than phonetically-arbitrary ones, assuming in both cases that the training data instantiates the patterns equally well, that the learner perceives the training data correctly, and that other factors are controlled for. By skewing the direction of language change, these biases could cause simple or phonetically-motivated patterns to accumulate, thus producing a corresponding skew in natural-language typology. (Of course, these inductive biases are not the only factors that could be affecting typological frequencies, and in particular substantive inductive bias is not the only way that phonetic substance could be influencing typology. More will be said of this in Part II of this paper.)

Natural-language typology is consistent with the effects of both kinds of bias, but the same is not true of phonological learning in the lab. Studies which directly compare simple patterns with complex ones nearly always find an advantage for the simple one, whereas studies which directly compare similar patterns instantiated by different features usually find no significant advantage for the phonetically-motivated or typologically-frequent pattern. Since similar methods and participant populations are used in both kinds of study, the systematic success with complexity and failure with substance corroborate the early conclusions of Pycha et al. (2003): If substantive biases exist at all, they are considerably weaker than complexity biases.[2]

[1] By *inductive bias* (also called "analytic bias" or "learning bias"), we mean any tendency of a pattern-learning algorithm to acquire one pattern faster or better than another from training sets that instantiate both patterns equally well. This definition is deliberately broad. It includes absolute distinctions between learnable and utterly unlearnable patterns, as well as relative distinctions between easier and harder ones. It is indifferent to details of implementation, applying to explicit penalties against specific patterns as well as to emergent consequences of the learner's architecture, representational system, or similarity metric. The term as we use it includes, but is not limited to, anything that would qualify as "Universal Grammar".

[2] The validity and necessity of considering null results (i.e., failures to find a statistically-significant difference) is discussed further in our companion paper (Part II, §7). For now, we note that studies which

The scope of this review is the effects of formal complexity (this paper) and phonetic substance (the companion paper, Part II) on the acquisition of artificial analogues of categorical phonology, i.e., patterns which partition a discrete stimulus space into positive ("legal", "pattern-conforming") versus negative ("illegal", "non-conforming") instances on the basis of phonological features.[3] In the non-linguistic psychological literature, such partitions are often referred to as "concepts". We are not concerned here with learning how to partition a continuous stimulus space into phonetic categories (e.g., Maye et al., 2002; Goudbeek et al., 2008), nor with analogues of lexical (e.g., Peña et al., 2002; Perruchet et al., 2004; Newport and Aslin, 2004) or syntactic (e.g., Gómez, 2002) dependencies. For a review of artificial-language research that extends outside phonology, see Culbertson 2012.

We have tried to be as inclusive as possible, but have necessarily omitted some studies, either because they do not fall within the purview of the article, or because their results were not reported in a way that we could use, or because a discussion would have consisted of criticizing the experiments rather than interpreting their results, or because we were simply unaware of them at the time when the article went to press. Our interpretations of the experiments which we do review are not necessarily those of the original authors.

2 Number of relevant features

Many formal theories of natural-language phonology are designed to favor patterns which have a simple expression in phonetic terms, and some proposals expressly impute this bias to human learners (e.g., Chomsky and Halle 1968, 330–334; Kiparsky 1971, 623; Bach and Harms 1972; Sagey 1990, 1; Hayes 1999; Gordon 2004, 304). Two main formal complexity factors have been studied in phonology: the number of features relevant to the pattern, and the relations between them. This section addresses the former.

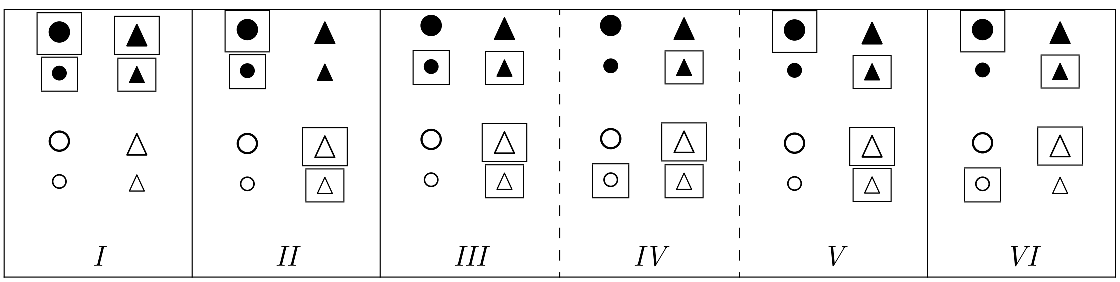
found null results *must* be reviewed because excluding them would falsely inflate the apparent robustness of the effects. Five percent of experiments testing a false hypothesis will find statistically significant support for it at the conventional $p < 0.05$ level. If the inconclusive or contradictory results which constitute the other 95% are suppressed on the grounds that they are "null", the hypothesis will wrongly appear to be true.

[3]By *feature*, we mean any discrete-valued variable created by partitioning a continuous or discrete physical dimension. We define a *phonetic feature* to me one for which the physical dimension is phonetic, and a *phonological feature* as a phonetic feature that is used in a model of human phonology. Experimental patterns are implemented in terms of phonetic features but interpreted in terms of phonological ones, with the result that the formal complexity and phonetic substance of a pattern depend on the choice of model. We will return to this point in our companion paper (Part II, §7.1) once individual examples have been presented.

As other authors (and a reviewer) have pointed out, natural-language phonological features are not always tied so closely to phonetic dimensions (Mielke, 2004). However, phonetically-arbitrary natural-language phonological features have not to our knowledge been studied in phonological learning experiments, so the above definitions are adequate for this review.

For an independently motivated standard of formal complexity, we turn to the psychology literature. A stimulus space described by three binary-valued features can be divided into two equal-sized categories in only six ways, if we ignore trivial variants obtained by permuting features or inverting feature values. Examples are shown in (1) for stimuli that are geometric figures varying in color (black vs. white), shape (circle vs. triangle), and size (large vs. small). Only color matters for the Type I distinction. Type II requires attention to color and shape, but size can be ignored. Types III through V involve all three features, but some subsets can be decided with fewer (e.g., white triangles). For Type VI, not even this is possible; even a subset requires all three features.

(1) *Representatives of the six possible equal partitions of a stimulus space defined by binary features of color, shape, and size. Boxes enclose the (arbitrary) positive class. Concepts are arranged in increasing order of difficulty, with III, IV, and V being about equal. (After Shepard et al. 1961.)*



These six concepts have been extensively studied in connection with supervised learning of non-linguistic categories.[4] In a typical experiment, the participant is shown a randomly-selected stimulus, judges whether it belongs to the target concept, and is then told the correct response. This cycle repeats until some performance criterion is met. The main finding is that difficulty increases along with the number of relevant features: Type I is easier than Type II, which is easier than Types III, IV, and V, which in turn are easier than Type VI (Shepard et al., 1961; Neisser and Weene, 1962; Nosofsky et al., 1994; Feldman, 2000; Love, 2002; Smith et al., 2004).

Analogous phonological stimuli have been used in both supervised and unsupervised learning experiments. Participants are either trained with feedback to divide stimuli into "legal" and "illegal" categories, or are familiarized without feedback on "legal" stimuli only. They are then asked to categorize stimuli as legal or illegal, or to decide which of two

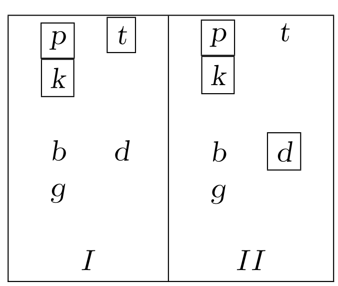
[4]The established terminology here invites confusion. The psychology literature uses "category" or "concept" to mean a partition of a stimulus space into disjoint labelled subsets. For example, "black triangles" would be a category or concept in the three-dimensional space shown in (1). In phonology and phonetics, "category" often has the more specific meaning of a phoneme, a region in phonetic space whose elements are phonologically equivalent (as in the phrase "categorical perception". We use the terms here in their broader sense; e.g., "voiced stops" would be called a category or concept.

stimuli is the legal one. The phonological experiments have replicated the non-linguistic difficulty hierarchy for Types I, II, and VI. We know of no published studies on the other three types.[5]

A phonological pattern that depends on a single stimulus feature (Type I) has often proven easier, and never harder, than one that requires more. Saffran and Thiessen (2003, Exps. 2, 3) familiarized English-learning 9-month-olds with isolated positive nonword instances, exposed them to a continuous stream of two positive and two negative nonwords, and then compared listening times to these four nonwords using headturn preference. When the pattern restricted [p t k] to some positions and [b d g] to others, the negative stimuli were preferred, but when the pattern distinguished [p d k] vs. [b t g], there was no difference in means (see schematic in Figure 2a). Cristiá and Seidl (2008, Exp. 1) familiarized English-learning 7-month-olds on positive C_1VC_2 nonwords. When the pattern was "C_1 is a nasal or oral stop" (i.e., [−continuant]), the infants preferred novel negative instances over positive ones; when it was "C_1 is a nasal or fricative", for which there is no standard feature, they showed no preference (Figure 2b). LaRiviere et al. (1974, 1977) trained English-L1 adults to categorize a set of six or eight syllables into two equal-sized classes defined either by a single feature or in an unspecified "random" way that needed more relevant features. Performance was significantly better for the single-feature condition than the random condition in three out of ten experiments, and was numerically better in the other seven.[6]

(2) *Examples of phonological patterns in experiments comparing featural complexity. Some are defective representatives of their category types, owing to the impossibility of some feature combinations. In each case, the left-hand pattern proved the easier. (Compare Table (1).)*

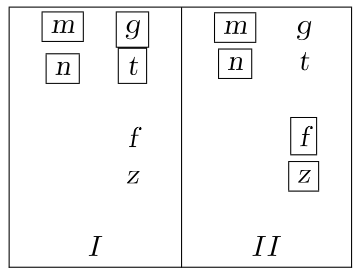
(a) Saffran and Thiessen (2003, Exps. 2, 3). Features were voiced vs. voiceless, coronal vs. non-coronal, and labial vs. velar. (Defective.)



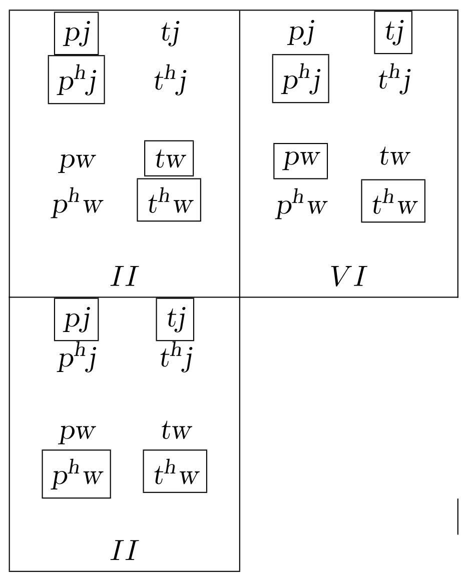
[5]There is an interesting discussion by Silverman 1999, 2006 of Shepard et al. (1961) in connection with the featural structure of a set of allophones of a phoneme, with particular attention to Shepard et al.'s findings about the effects of practice in reducing the differences in difficulty between types. These papers fall outside the scope of this review since they do not present phonological learning data.

[6]If there were in fact no difference between the simple and complex conditions in any of the experiments, then the chance that all ten experiments would favor the simple condition is 1 in 1024, or 0.000977.

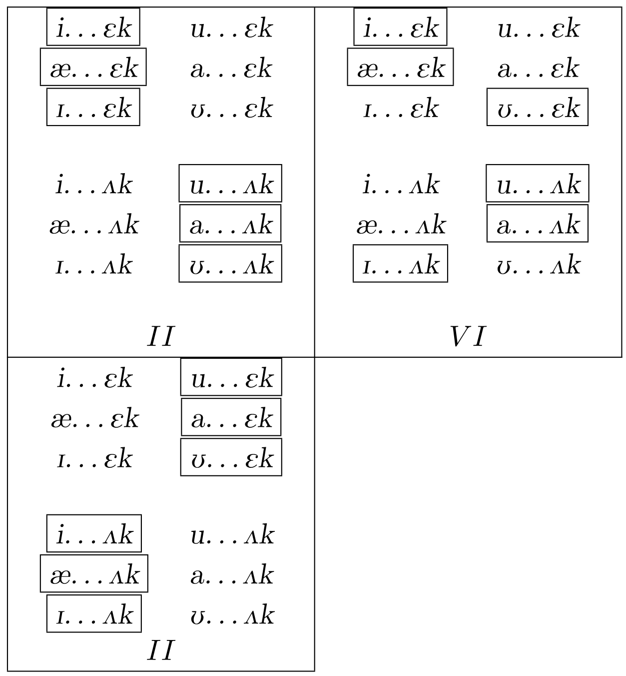
(b) *Cristiá and Seidl (2008): Features were oral vs. nasal, continuant vs. non-continuant, labial vs. non-labial. (Defective.)*



(c) *Kuo (2009): Features were plain initial stop vs. aspirated, labial initial stop vs. coronal, and palatal glide vs. labiovelar. Corresponding conditions (not shown) inverted the legal/illegal categories.*



(d) Pycha et al. (2003): Features are front first vowel vs. back, front last vowel vs. back, and high-lax first vowel vs. other first vowel.



Two-feature Type II patterns enjoy a similar advantage over three-feature Type VI ones. Kuo (2009) familiarized L1 Mandarin speakers on syllables with two-consonant onsets in which the initial stop perfectly predicted whether the following glide would be [j] or [w] (Figure 2c). In two patterns, a single stop feature, aspiration or place, was relevant (Type II); in the third, both stop features were needed (Type VI). Both Type II conditions elicited a significantly greater preference for novel positive stimuli over non-conforming foils than did the Type VI condition. A similar result was found by Pycha et al. (2003) in an experiment in which adult English speakers were trained with feedback to make binary grammaticality judgments of $X \ldots XY$ stimuli, where Y was either [-ɛk] or [-ʌk] depending on the vowel of X (Figure 2d). Classification at test was more accurate for two Type II patterns (backness agreement and backness disagreement) than for a Type VI pattern. Skoruppa and Peperkamp (2011) exposed French speakers to spoken passages in their own language which had been modified so that a front vowel either agreed in rounding with the preceding vowel (Type II), disagreed, (Type II), or agreed if mid and disagreed if high (Type VI). Participants in the Type II conditions were better at recognizing new pattern-conforming stimuli than those in the Type VI condition.

The advantage for patterns with fewer relevant features extends to patterns which are in part phonetically arbitrary. Using a speeded-repetition paradigm, Chambers et al. (2010) familiarized English speakers with a pattern in which the unsystematic sets [b f k m t] and [p s g n tʃ] were restricted to opposite ends of a C_1VC_2 syllable when the nucleus was one of two vowels, but unrestricted when it was a third vowel. In four experiments with different vowel sets, novel probe syllables were repeated faster when their consonants obeyed the positional restriction, regardless of what the vowel was. Participants evidently did not detect the dependency between V and the presence of positional restrictions on the Cs. Their performance followed an inaccurate moderately-complex generalization rather than an accurate more-complex one.

A complexity disadvantage has also been reported for learned alternations in production. Peperkamp et al. (2006) exposed French-speaking adults to stimuli of the form XY paired with pictures of two or three of the same object. The number of objects determined X ([nɛl] or [ʀa]), and the identity of the object determined Y. The initial consonant C_Y of each Y varied depending on X. In two conditions, a phonetically-defined set of consonants switched voicing ([f ʃ]↔[v ʒ], or [p k]↔[b g]). In two others, both the sets and the change were phonetically unsystematic ([p z]↔[ʒ f], or [ʃ v]↔[b k]). When tested on XY phrases with novel Ys, participants in the voicing conditions changed C_Y in the pattern-conforming way about 25% of the time, whereas those in the unsystematic conditions did so only about 5% of the time (most responses left C_Y unchanged). The relevant complexity here seems to be that of the change rather than that of the segment classes undergoing it: Participants in the voicing conditions did not generalize the rule to new segments in the old classes. They must have induced two single-segment rules rather than a class-based one (e.g., [f]↔[v] and [ʃ]↔[ʒ], rather than "voiceless fricatives alternate with voiced ones"). Evidently a rule like [f]↔[v] is learned better than one like [ʃ]↔[b]. Using similar stimuli with a similar population, Skoruppa et al. (2009) found that an alternation in which only place of articulation changed was learned better than one in which place and manner changed together, and also better than another in which place, manner, and voice changed together.

On the other hand, an experiment by Finley and Badecker (2010) did not find a preference for a two-feature change over a one-feature change. They familiarized English speakers using stimuli consisting of three syllables X, Y, z, where X and Y agreed in backness but disagreed with z, followed by their concatenation XYZ with the final syllable harmonized to the first two (i.e., every trial changed only the third syllable).[7] At test, they strongly preferred the familiar A, B, c, ABC, with one change, over A, B, c, abc, with two, but the one-change preference was reduced or even slightly reversed when the choice was between the unfamiliar stimulus types A, b, c, ABC (two changes) and A, b, c, abc (one). However, any bias towards fewer changes would have to compete with an opposing familiarity difference: A, b, c, ABC has an unfamiliar change in the second syllable; A, b, c, abc, an unfamiliar change in the first syllable and an unfamiliar lack of change in the third syllable. (Unfamiliarized participants had no significant preference.)

Thus, phonological-learning results from a wide variety of paradigms converge to show

[7] A mirror-image condition reversed the role of the first and third syllables, without major changes in the outcome.

that patterns become harder to learn as the number of relevant features increases. These results are in agreement with what has been found for the learning of non-linguistic patterns.

3 Relations between features

A separate question is whether certain syntagmatic relations between features *within* a stimulus facilitate pattern learing when the number of relevant features is controlled. Two main relations have been studied in artificial phonology, featural agreement and the contiguity-similarity tradeoff.

3.1 Featural agreement

Dependencies between instances of the same feature within a word are conspicuously common cross-linguistically in the form of assimilation and dissimilation patterns (Archangeli and Pulleyblank, 2011; Bakovć, 2011; Bye, 2011; Rose and Walker, 2011). Many phonological theories make special provision for representing agreement or disagreement of features within some part of an utterance (e.g., Chomsky and Halle, 1968; Goldsmith, 1976; Alderete and Frisch, 2008). Domain-general theories of category learning have not addressed this possibility, and lack the means to recognize two features in the same stimulus as instances of the same abstract feature (Gluck and Bower, 1988; Kruschke, 1992; Nosofsky et al., 1994; Love et al., 2004; Feldman, 2006). However, there is non-linguistic evidence that patterns are easier to learn when they relate, e.g., color to color, or shape to shape, than when they relate values on two different dimensions (Hunt and Hovland 1960; Ciborowski and Cole 1973; Ciborowski and Price-Williams 1974; not found by Shepard et al. 1961). This is true even for intra-dimensional patterns other than agreement (Rogers and Johnson, 1973). Several studies of phonological learning have found a similar phenomenon: Patterns relating two instances of the same feature produce larger familiarity effects than those relating instances of two different features.

Wilson (2003) familiarized English speakers to stimuli of the form $C_1V_1C_2V_2C_3a$. The identity of C_2 determined whether C_3 was [n] or [l], as shown in (3).

(3) *Conditions of Experiments 1 and 2 of Wilson (2003).*

	Familiarization			*% judged familiar*			
	Nasal C_2	*Dorsal* C_2	*Other* C_2	*Old*		*Novel*	
Exp.	4×	4×	12×	*Conf.*	*Non.*	*Conf.*	*Non.*
1A	$C_1V_1[m/n]V_2na$	$C_1V_1[k/g]V_2la$	$C_1V_1C_2V_2la$	*70*	*44*	*53*	*34*
1B	$C_1V_1[m/n]V_2la$	$C_1V_1[k/g]V_2na$	$C_1V_1C_2V_2la$	*60*	*54*	*46*	*38*
2A	$C_1V_1[m/n]V_2la$	$C_1V_1[k/g]V_2na$	$C_1V_1C_2V_2na$	*73*	*47*	*50*	*35*
2B	$C_1V_1[m/n]V_2na$	$C_1V_1[k/g]V_2la$	$C_1V_1C_2V_2na$	*68*	*52*	*47*	*41*

The patterns in Conditions 1A and 2A can be stated as agreement or disagreement in [nasal] between C_2 and C_3, whereas those in 1B and 2B crucially involve a relation between two features [Dorsal] and [nasal], or [Dorsal] and [lateral]. Concept membership significantly increased judged familiarity in the single-feature Conditions 1A and 2A, but not in the two-feature Conditions 1B and 2B.[8] ,[9]

Healy and Levitt (1980, Experiment 3) found that a voicing-conditioned pattern was acquired better by English speakers than a phonetically arbitrary one when the pattern was voicing assimilation, but not when it was a correlation between voicing and vowel quality ([a] vs. [o]).[10] Moreton (2008, 2012) familiarized English speakers on $C_1V_1C_2V_2$ stimuli and tested discrimination between novel positive and negative instances. Performance was better when the pattern was height agreement between the vowels, or voice agreement between the consonants, than when it was correlation between the height of V_1 and voicing of C_2. This phenomenon is not peculiar to English, as Lin (2009) found the same result with speakers of Mandarin and speakers of Southern Min using a similar paradigm and the same stimuli as Moreton (2008); nor is it confined to vowel height, since performance was better for backness agreement between the vowels than for correlation between the backness of V_1 and the voicing of C_2 (Moreton, 2012). Height-backness and voice-place dependencies showed no such advantage over height-voice, indicating a specific advantage for dependencies that involve two instances of the same feature, over and above any more general advantage for consonant-to-consonant or vowel-to-vowel dependencies.

Not all experiments that could have detected a single-feature advantage actually did so. Kuo (2009) found no difference between a place-place correlation (labial glide iff labial stop) and a place-aspiration correlation (labial glide iff aspirated stop; see the two Type II

[8]Conditions 1A and 2A partially confound featural identity with segmental identity in the case of [n]; however, the difference between the C_2 = [m] and C_2 = [n] sub-conditions was not significant (Colin Wilson, p.c., 2010).

[9]This interpretation hinges on the traditional assumption that English post-tonic intervocalic [l] is phonologically [Coronal] despite its phonetic dorsal component (Sproat and Fujimura, 1993, 304). If [l] is phonologically [Dorsal] as well (Walsh Dickey, 1997, Ch. 2), then all four conditions can be stated as single-feature agreement or disagreement, and this study may exemplify a substantive rather than a complexity bias.

[10]The "arbitrary" patterns were constructed in the same way as those of Saffran and Thiessen (2003), by replacing the classes [p t k] vs. [b d g] with [p t g] vs. [b d k].

patterns in Figure 2c). Jennifer Smith points out to us that in some proposed phonological-feature systems, consonant labiality and vowel labiality are two different features (Odden, 1991), making both patterns two-feature patterns. Another null result was found by Seidl and Buckley (2005, Experiment 2), who familiarized 9-month-old infants on $C_1V_1C_2V_2(C_3)$ stimuli and tested listening preference for novel positive vs. negative stimuli. A novelty preference was obtained for an agreement pattern in which C_1 and V_1 agreed in labiality, but also for one in which C_1 was labial if and only if V_1 was high.

The overall tenor of the evidence is that dependencies between two instances of the same feature produce larger familiarity effects than dependencies between instances of two different features. These aspects of phonological learning have analogues in non-linguistic learning (reviewed in Moreton 2012).

3.2 Contiguity-similarity tradeoff

Phonological theory typically treats dependencies between adjacent elements as the normal case, excluding long-distance interactions unless the interacting segments share some property which is absent from intervening material (Jensen, 1974; McCarthy, 1981; Cole and Trigo, 1988; Pierrehumbert, 1993; Odden, 1995; Gafos, 1996; Hansson, 2001; Frisch et al., 2004; Rose and Walker, 2004; Heinz, 2010). As with the other formal complexity biases reviewed above, there are parallels in non-linguistic learning: Two stimuli, or two elements of a compound stimulus, are more likely to cohere in perception and become associated in memory if they are contiguous in time or space, or are perceptually similar (Köhler 1941; Prentice and Asch 1958; Asch 1969; Arnold and Bower 1972; Rescorla 1980; Rescorla and Gillan 1980; Creel et al. 2004; Rescorla 2008; but see Pacton and Perruchet 2008.) It would therefore be surprising if contiguity and similarity did not facilitate acquisition within-stimulus dependencies in the lab. However, the relevant evidence is scanty, and what there is of it does not indicate a strong effect of either factor.[11]

Shorter-range dependencies have proven little, if any, easier to learn than longer-range ones. Majerus et al. (2004) familiarized French speakers on a continuous stream of CV syllables which contained phonetically unsystematic CV and $C \dots C$ dependencies, then tested immediate recall of novel isolated nonword probes. Probes which belonged to both patterns simultaneously were recalled better than those which belonged to neither or only one, but there was no evidence that the two patterns differed in effect. Using a tongue-twister paradigm, Warker and Dell (2006, 2008) tested English speakers on stimuli in which two consonants ([f] and [s], or [k] and [g]) were constrained to appear at opposite ends of the stimulus. The positional restrictions were reversed depending on a third segment which was either adjacent to the marginal consonants ($C_1\mathbf{V}C_2$) or remote from them ($C_1V\mathbf{C}VC_2$). Exchange errors followed the positional restrictions to almost the same ex-

[11] We omit here studies of artificial long-distance lexical (e.g. Peña et al., 2002; Perruchet et al., 2004; Newport and Aslin, 2004) and syntactic (e.g. Gómez, 2002) dependencies, as outside the scope of this review.

tent in the adjacent and remote conditions, with a numerical but nonsignificant advantage for the adjacent condition. Koo and Callahan (2011) used $C_1V_1C_2V_2C_3V_3$ stimuli to familiarize English speakers on a phonetically arbitrary $C_1 - C_2$ or $C_1 - C_3$ dependency. In both pattern conditions, novel conforming test items were judged as familiar significantly more often than were novel nonconforming items, and there were no significant differences between the closer and more-distant dependencies. Moreton (2008, 2012) familiarized English speakers on $C_1V_1C_2V_2$ stimuli and then tested their ability to discriminate novel positive versus negative stimuli. Conformity preference for phonetically-systematic continguous $V_1 - C_2$ dependencies did not differ significantly from that for phonetically-systematic non-contiguous $C_1 - V_2$, $C_1 - C_2$, or $V_1 - V_2$ dependencies, except when the latter involved two instances of the same feature (e.g., height-height)—in which case the stronger effect was produced by the *non*-contiguous dependency.

A similar result was obtained by Finley (2011a), but with an interesting twist. English-speaking participants were familiarized on stimuli of the form $X \ldots XY$, where X always contained exactly one [s] or [ʃ], and Y was [su] or [ʃu] to match. They were then tested on their ability to choose the pattern-conforming member of an X[su]–X[ʃu] pair. When all the familiarization stimuli had the form CV[s/ʃ]V[su/ʃu], with the critical consonants in adjacent syllables, participants in the test phase preferred new stimuli that fit the pattern over those that did not. Familiarization on [s/ʃ]VCV[su/ʃu], with the critical consonants far apart, likewise led to a preference for conforming test stimuli. As in other studies, there was no significant difference between the shorter- and the longer-range dependency in the rate of pattern-conforming responses. However, participants familiarized on [s/ʃ]VCV[su/ʃu] also significantly preferred CV[s/ʃ]V[su/ʃu] over CV[s/ʃ]V[ʃu/su], whereas those familiarized on the shorter-range dependency did not show an analogous preference for [s/ʃ]VCV[su/ʃu]-conforming test stimuli. Thus, familiarization on the longer-range pattern generalized to the shorter-range one, but not vice versa.[12] Subsequent experiments using $C_1V_1C_2V_2C_3V_3$[su/ʃu] stimuli found that a familiarized identity relation between C_1 and C_4 was learned as well as one between C_2 and C_4, and that each generalized equally well to the other (Finley, 2011b).

In none of these studies have local dependencies produced a greater conformity preference than remote ones. Although no experiment directly tested for an interaction between contiguity and similarity, results have been comparable across experiments regardless of whether the segments involved are very similar, like the [s] and [ʃ] of Finley (2011a), or very different, like the [s] and [m] of Koo and Callahan (2011). The resemblance to natural-language typology, where local dependencies are the norm, is rather weak, raising the possibility that the typological bias may not be wholly due to inductive bias.

[12] This result is somewhat qualified by the fact that, when the rates of pattern-conforming response in the two generalization conditions were compared directly, the difference was only marginally significant.

4 Summary: Complexity

There is abundant converging evidence that formal complexity impedes acquisition of artificial phonological patterns, in the sense that performance drops as the number of relevant features increases. The strongest result is that Type I patterns are easier than Type II, which are easier than Type VI (see §2 and Table 1). There is also evidence that stimulus-response mappings which change fewer features are easier than those which change more, and that within-stimulus dependencies are easier when they involve two instances of the same feature (i.e., assimilation or dissimilation) than instances of two different features. If the same inductive biases affect natural-language phonology, they should leave visible marks on cross-linguistic typology and within-language productivity.

A complexity bias in natural-language phonology would make more complex patterns harder to learn, hence harder to innovate and more likely to be changed (simplified) in transmission from one generation to the next (Bach and Harms, 1972). That in turn could lead to low long-term steady-state frequencies for the corresponding patterns (Bell, 1970, 1971; Greenberg, 1978). Phonologists have in fact noted informally that the patterns they discover tend to be featurally simple (e.g., Chomsky and Halle 1968, 401, Hayes 1999, Pierrehumbert 2001). The available quantitative evidence tends to confirm this observation.

Clements (2003) found that inventories tend to avoid both "holes" and "bumps": A given segment is more likely if all of its feature values are shared by other segments, and less likely if some of them are not. He proposes that inventories tend to maximize *feature economy*, the ratio of the number of segments in an inventory to the number of features required to distinguish among inventory members. Feature economy favors Shepard Type I inventories over those of Types III—VI. An example is shown in Figure (4). All three inventories contain four sounds, but the Type I inventories use only two contrastive features (feature economy index = 4/2 = 2), while the Type V inventory uses three (feature economy index = 4/3).

(4) *The probability that a segment will occur in an inventory increases if the inventory contains other segments minimally different from it. (Extrapolated from Clements, 2003, Figure 11).*

Favored		*Disfavored*
[*p*] [*t*]	[*p*] [*t*]	[*p*] [*t*]
**b* **d*	[*b*] [*d*]	[*b*] **d*
[*f*] [*s*]	**f* **s*	**f* **s*
**v* **z*	**v* **z*	**v* [*z*]
Type I	*Type I*	*Type V*

In a survey of 561 languages, Mielke (2004, 2008) studied "phonologically active classes", sets of sounds that pattern together by undergoing an alternation, triggering an alternation, or respecting a phonotactic restriction. One finding was that typologically common sound classes can usually be stated as a single feature value or a conjunction of a small number of feature values (.e.g, [–continuant –sonorant]), with typological frequency falling as the feature count rises. Of the non-conjunctive classes, most can be stated as disjunctions of conjunctions (e.g., [–sonorant Labial]∨[–sonorant Dorsal]); frequency falls as the number of disjuncts rises. Thus, featurally-complex patterns are attested but rarer, in the same way in which their artificial analogues are learnable but harder.[13]

An alternative hypothesis attributes the prevalence of simple classes to sampling error Pierrehumbert (2001, 2003). Suppose the learner decides whether to postulate a constraint by observing which of two classes, A or B, is more frequent. Classes defined by more features are rarer ($[+F+G]$ cannot outnumber $[+F]$), so if A and B both involve many features, the corpus of relevant examples will be small. For example, the learner can make a more reliable frequency comparison between $A = \{$ plosives $\}$ and $B = \{$ fricatives $\}$ than between $A' = \{$ labial plosives $\}$ and $B' = \{$ labial fricatives $\}$. Since there is more variability between smaller samples, learners will disagree more in their judgment of the relative frequency of A' and B', and hence also in the constraints they acquire. That makes highly specific ("fine-grained") constraints less likely to survive inter-generational transmission than very general ones. This idea may explain why conjunctive categories with more features are typologically rarer. However, it does not explain why *dis*junctive classes become rarer as the number of disjuncts goes up, since more disjuncts mean a larger class: The learner can make an even more reliable frequency comparison between $A^* = \{$ plosives and nasals $\}$ and $B^* = \{$ fricatives and laterals $\}$ than between A and B. In contrast, a bias towards featural simplicity can account equally well for the rarity of both complex conjunctions and complex disjunctions.

Another alternative is that the simplicity of natural-language phonology is inherited from the simplicity of natural-language phonetics. The phonological form of an utterance perceived by a listener is sometimes different from that intended by the speaker, owing to systematic distortions introduced by the articulation-transmission-perception channel. Such channel biases may serve as phonetic precursors for phonological innovations, e.g., if phonetic coarticulation by speakers is interpreted as phonological assimilation by listeners ("phonologization", Hyman 1976; Ohala 1993). If the precursors tend to be simple in phonetic terms, then their phonologizations will tend to be simple in phonological terms.

This hypothesis is not yet testable because there is no Mielke (2008)-like quantitative data on the complexity of phonetic precursors. In the meantime, it is more feasible to test a stronger hypothesis, namely, that phonological patterns tend to be simple *only*

[13] Many of the natural-language classes in the Mielke (2004) study involve more relevant features than the artificial-phonology experiments. However, they agree where they overlap. A Type I problem uses a single-feature class. A Type II problem requires a disjunction of two two-feature conjunctions, e.g., "(black and triangle) or (not-black and non-triangle)", while Type VI needs four three-feature disjuncts (Feldman, 2000, 2006; Lafond et al., 2007).

because precursors do (and not because inductive bias also favors innovation or preservation of simpler phonological patterns). Under this hypothesis, precursors which are equal in phonetic magnitude and differ only in complexity should be phonologized at the same rate, yield equally durable phonological patterns, and hence have the same typological frequency. On the other hand, if typology is also shaped by an inductive bias towards simplicity, the simpler precursor should be phonologized more often, yield a more-durable phonological pattern, and thus accumulate greater typological frequency.

Two such pairs of precursors have been examined in connection with a hypothesized inductive bias favoring single-feature dependencies over two-feature ones. (1) Vowel-height harmony and disharmony are typologically more common than phonological patterns relating vowel height to consonant voicing. However, the apparent phonetic precursors of these two patterns do not differ in their effect on first-formant frequency (Moreton, 2008). (2) Phonological patterns relating the height of tones in adjacent syllables are typologically more common than those relating tone height to consonant voicing, yet phonetic tonal coarticulation and tone-voice interaction do not differ in their effect on f_0 (Moreton, 2010). Both of these results suggest that the typological skew towards simplicity is at least partly due to inductive bias. They are on the one hand weakened by the use of acoustic rather than perceptual measures of phonetic-precursor magnitude (Yu, 2010, 2012), but on the other hand strengthened by evidence of an analogous inductive bias in learning experiments with English, Mandarin, and Southern Min speakers (see above, §3.1, p. 10).

We have seen that the difficulty of a pattern is strongly affected by the abstract structural relationships between the more-primitive features that define it. We now turn in our companion article (Part II) to the question of whether the phonetic content of those features matters as well.

References

Alderete, J. and S. A. Frisch (2008). Dissimilation in grammar and the lexicon. In P. de Lacy (Ed.), *The Cambridge handbook of phonology*, Chapter 16, pp. 379–398. Cambridge, England: Cambridge University Press.

Archangeli, D. and D. Pulleyblank (2011). Harmony. In P. de Lacy (Ed.), *The Cambridge handbook of phonology*, Chapter 15, pp. 353–378. Oxford, England: Cambridge University Press.

Arnold, P. G. and G. H. Bower (1972). Perceptual conditions affecting ease of association. *Journal of Experimental Psychology 93*(1), 176–180.

Asch, S. E. (1969). A reformulation of the problem of associations. *American Psychologist 24*(2), 92–101.

Bach, E. and R. T. Harms (1972). How do languages get crazy rules? In R. P. Stockwell and R. K. S. Macaulay (Eds.), *Linguistic change and generative theory*, Chapter 1, pp. 1–21. Bloomington: Indiana University Press.

Bakovć, E. (2011). Local assimilation and constraint interaction. In P. de Lacy (Ed.), *The Cambridge handbook of phonology*, Chapter 14, pp. 335–352. Oxford, England: Cambridge University Press.

Bell, A. (1970). *A state-process approach to syllabicity and syllabic structure.* Ph. D. thesis, Stanford University.

Bell, A. (1971). Some patterns of the occurrence and formation of syllabic structure. *Working Papers on Language Universals 6*, 23–138.

Bye, P. (2011). Dissimilation. In M. van Oostendorp, C. Ewen, and E. Hume (Eds.), *The Blackwell companion to phonology*, Chapter 63. Wiley-Blackwell.

Chambers, K. E., K. H. Onishi, and C. Fisher (2010). A vowel is a vowel: generalizing newly learned phonotactic constraints to new contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition 36*(3), 821–828.

Chomsky, N. and M. A. Halle (1968). *The sound pattern of English.* Cambridge, Massachusetts: MIT Press.

Ciborowski, T. and M. Cole (1973). A developmental and cross-cultural study of the influences of rule structure and problem composition on the learning of conceptual classifications. *Journal of Experimental Child Psychology 15*(2), 193–215.

Ciborowski, T. and D. Price-Williams (1974). The influence of rule structure and problem composition on conceptual learning among rural Hawaiian children. Technical Report 75, Kamehameha Early Education Program, Kamehamea Schools and Bernice P. Bishop Estate, 1850 Makuakane Street, Honolulu, Hawaii 96817.

Clements, G. N. (2003). Feature economy in sound systems. *Phonology 20*(3), 287–333.

Cole, J. and L. Trigo (1988). Parasitic harmony. In H. van der Hulst and N. Smith (Eds.), *Features, segmental structure, and harmony processes*, pp. 19–38. Foris.

Creel, S. C., E. L. Newport, and R. N. Aslin (2004). Distant melodies: statistical learning of non-adjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition 30*, 1119–1130.

Cristiá, A. and A. Seidl (2008). Is infants' learning of sound patterns constrained by phonological features? *Language Learning and Development 4*(3), 203–227.

Culbertson, J. (2012). Typological universals as reflections of biased learning: evidence from artificial language learning. *Language and Linguistics Compass 6*(5), 310–329.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature 407*, 630–633.

Feldman, J. (2006). An algebra of human concept learning. *Journal of mathematical psychology 50*, 339–368.

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

Finley, S. (2011b). Testing the limits of long-distance learning: learning beyond the three-segment window. To appear in *C*ognitive Science.

Finley, S. and W. Badecker (2010). Linguistic and non-linguistic influences on learning biases for vowel harmony. In S. Ohlsson and R. Catrambone (Eds.), *Proceedings of the*

32nd Annual Conference of the Cognitive Science Sciety, Austin, Texas, pp. 706–711. Cognitive Science Society.

Frisch, S., J. B. Pierrehumbert, and M. B. Broe (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory 22*(1), 179–228.

Gafos, A. (1996). *The articulatory basis of locality in phonology.* Ph. D. thesis, Johns Hopkins University, Baltimore.

Gallistel, C. R., A. L. Brown, S. Carey, R. Gelman, and F. C. Keil (1991). Lessons from animal learning for the study of cognitive development. In S. G. Carey and R. G. Gelman (Eds.), *The epigenesis of mind*, Chapter 1, pp. 3–36. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Gluck, M. A. and G. H. Bower (1988). Evaluating an adaptive network model of human learning. *Journal of Memory and Language 27*, 166–195.

Goldsmith, J. A. (1976). *Autosegmental phonology.* Ph. D. thesis, Massachusetts Institute of Technology.

Gómez, R. (2002). Variability and detection of invariant structure. *Psychological Science 13*(5), 431–436.

Gordon, M. (2004). Syllable weight. In B. Hayes, R. Kirchner, and D. Steriade (Eds.), *Phonetically-based phonology*, pp. 277–312. Cambridge, England: Cambridge University Press.

Goudbeek, M., A. Cutler, and R. Smits (2008). Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech communication 50*, 109–125.

Greenberg, J. H. (1978). Diachrony, synchrony, and language universals. In J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (Eds.), *Universals of human language, volume 1, method and theory*, pp. 61–91. Stanford, California: Stanford University Press.

Hansson, G. Ó. (2001). The phonologization of production constraints: Evidence from consonant harmony. In M. Andronis, C. Ball, H. Elston, and S. Neuvel (Eds.), *Papers from the 37th meeting of the Chicago Linguistics Society (CLS 37)*, Volume 1, Chicago, pp. 187–200. Chicago Linguistic Society.

Hayes, B. (1999). Phonetically driven phonology: the role of optimality in inductive grounding. In M. Darnell, E. Moravcsik, M. Noonan, F. Newmeyer, and K. Wheatly (Eds.), *Functionalism and Formalism in Linguistics*, Volume 1: General Papers, pp. 243–285. Amsterdam: John Benjamins.

Healy, A. F. and A. G. Levitt (1980). Accessibility of the voicing distinction for learning phonological rules. *Memory and Cognition 8*(2), 107–114.

Heinz, J. (2010). Learning long-distance phonotactics. *Linguistic Inquiry 41*(4), 623–661.

Hunt, E. B. and C. L. Hovland (1960). Order of consideration of different types of concepts. *Journal of Experimental Psychology 59*(4), 220–225.

Hyman, L. M. (1976). Phonologization. In A. Juilland (Ed.), *Linguistic studies offered to Joseph Greenberg: second volume: phonology*, pp. 407–418. Saratoga, California: Anma Libri.

Jensen, J. T. (1974). A constraint on variables in phonology. *Language 50*(4, Part 1), 675–686.

Kiparsky, P. (1971). Historical linguistics. In W. Dingwall (Ed.), *A survey of linguistic science*, pp. 576–653. College Park: University of Maryland.

Köhler, W. (1941). On the nature of associations. *Proceedings of the American Philosophical Society 84*, 489–502.

Koo, H. and L. Callahan (2011). Tier adjacency is not a necessary condition for learning non-adjacent dependencies. *Language and Cognitive Processes*.

Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review 99*, 22–44.

Kuo, L. (2009). The role of natural class features in the acquisition of phonotactic regularities. *Journal of psycholinguistic research 38*(2), 129–150.

Lafond, D., Y. Lacouture, and G. Mineau (2007). Complexity minimization in rule-based category learning: revising the catalog of Boolean concepts and evidence for non-minimal rules. *Journal of Mathematical Psychology 51*, 57–75.

LaRiviere, C., H. Winitz, J. Reeds, and E. Herriman (1974). The conceptual reality of selected distinctive features. *Journal of Speech and Hearing Research 17*(1), 122–133.

LaRiviere, C., H. Winitz, J. Reeds, and E. Herriman (1977). Erratum: The conceptual reality of selected distinctive features. *Journal of Speech and Hearing Research 20*(4), 817.

Lin, Y. (2009). Tests of analytic bias in native Mandarin speakers and native Southern Min speakers. In Y. Xiao (Ed.), *21st North American Conference on Chinese Linguistics*, Smithfield, Rhode Island, pp. 81–92. Bryant University.

Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin and Review 9*(4), 829–835.

Love, B. C., D. L. Medin, and T. M. Gureckis (2004). SUSTAIN: a network model of category learning. *Psychological Review 111*(2), 309–332.

Majerus, S., L. Mulder, T. Meulmans, and F. Peters (2004). Verbal short-term memory reflects the sublexical organization of the phonological language network: evidence from an incidental phonotactic learning paradigm. *Journal of Memory and Language 51*, 297–306.

Maye, J., J. F. Werker, and L. Gerken (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition 82*, B101–B111.

McCarthy, J. J. (1981). A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry 12*, 373–418.

Mielke, J. (2004). *The emergence of distinctive features*. Ph. D. thesis, Ohio State University.

Mielke, J. (2008). *The emergence of distinctive features*. Oxford, England: Oxford University Press.

Mitchell, T. M. (1980 [1990]). The need for biases in learning generalizations. In J. W. Shavlik and T. G. Detterich (Eds.), *Readings in machine learning*, Chapter 2.4.1, pp.

184–191. San Mateo, California: Morgan Kaufman.
Moreton, E. (2008). Analytic bias and phonological typology. *Phonology 25*(1), 83–127.
Moreton, E. (2010). Underphonologization and modularity bias. In S. Parker (Ed.), *Phonological argumentation: essays on evidence and motivation*, Chapter 3, pp. 79–101. London: Equinox.
Moreton, E. (2012). Inter- and intra-dimensional dependencies in implicit phonotactic learning. *Journal of Memory and Language 67*(1), 165–183.
Neisser, U. and P. Weene (1962). Hierarchies in concept attainment. *Journal of Experimental Psychology 64*(6), 640–645.
Newport, E. and R. N. Aslin (2004). Learning at a distance i: statistical learning of non-adjacent dependencies. *Cognitive Psychology 48*, 127–162.
Nosofsky, R. M., M. A. Gluck, T. J. Palmeri, S. C. McKinley, and P. Gauthier (1994). Comparing models of rule-based classification learning: a replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory and Cognition 22*(3), 352–369.
Nosofsky, R. M., T. J. Palmeri, and S. C. McKinley (1994). Rule-plus-exception model of classification learning. *Psychological Review 101*(1), 53–79.
Odden, D. (1991). Vowel geometry. *Phonology 8*, 261–289.
Odden, D. (1995). Adjacency parameters in phonology. *Language 70*(2), 289–330.
Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: problems and perspectives*, pp. 237–278. Harlow: Longman.
Pacton, S. and P. Perruchet (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition 34*(1), 80–96.
Peña, M., L. L. Bonatti, M. Nespor, and J. Mehler (2002). Signal-driven computations in speech processing. *Science 298*, 604–607.
Peperkamp, S., K. Skoruppa, and E. Dupoux (2006). The role of phonetic naturalness in phonological rule acquisition. In D. Bamman, T. Magnitskaia, and C. Zoller (Eds.), *Papers from the 30th Boston University Conference on Language Development (BUCLD 30)*, pp. 464–475.
Perruchet, P., M. D. Tyler, N. Galland, and R. Peereman (2004). Learning nonadjacent dependencies: no need for algebraic-like computations. *Journal of Experimental Psychology: General 133*(4), 573–583.
Pierrehumbert, J. B. (1993). Dissimilarity in the Arabic verbal roots. In A. Schafer (Ed.), *Proceedings of the Northeast Linguistics Society*, Volume 23, Amherst, Mass., pp. 367–381. Graduate Linguistics Students Association.
Pierrehumbert, J. B. (2001). Why phonological constraints are so coarse-grained. *Language and Cognitive Processes 16*(5–6), 691–698.
Pierrehumbert, J. B. (2003). Probabilistic phonology: discrimination and robustness. In R. Bod, J. Hay, and S. Jannedy (Eds.), *Probabilistic phonology*, Chapter 6, pp. 177–228. Cambridge, Massachusetts: MIT Press.
Pinker, S. (1979). Formal models of language learning. *Cognition 7*, 217–283.

Prentice, W. C. H. and S. E. Asch (1958). Paired associations with related and unrelated pairs of nonsense-figures. *American Journal of Psychology 71*, 247–254.

Pycha, A., P. Nowak, E. Shin, and R. Shosted (2003). Phonological rule-learing and its implications for a theory of vowel harmony. In M. Tsujimura and G. Garding (Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics (WCCFL 22)*, pp. 101–114.

Rescorla, R. A. (1980). Simultaneous and successive associations in sensory preconditioning. *Journal of Experimental Psychology: Animal Behavior Processes 6*(3), 207–216.

Rescorla, R. A. (2008). Evaluating conditioning of related and unrelated stimuli using a compound test. *Learning and Behavior 36*(2), 67–74.

Rescorla, R. A. and D. J. Gillan (1980). An analysis of the facilitative effect of similarity on second-order conditioning. *Journal of Experimental Psychology: Animal Behavior Processes 6*(4), 339–351.

Rogers, C. J. and P. J. Johnson (1973). Attribute identification in children as a function of stimulus dimensionality. *Journal of Experimental Child Psychology 15*(2), 216–221.

Rose, S. and R. Walker (2004). A typology of consonant agreement as correspondence. *Language 80*(3), 475–531.

Rose, S. and R. Walker (2011). Harmony systems. In J. Goldsmith, J. Riggle, and A. C. L. Yu (Eds.), *The handbook of phonological theory* (2nd ed.)., Chapter 8, pp. 240–290. Malden, Massachusetts: Blackwell.

Saffran, J. R. and E. D. Thiessen (2003). Pattern induction by infant language learners. *Developmental Psychology 39*(3), 484–494.

Sagey, E. (1990). *The representation of features in non-linear phonology: the Articulator Node Hierarchy.* New York: Garland.

Seidl, A. and E. Buckley (2005). On the learning of arbitrary phonological rules. *Language Learning and Development 1*(3 & 4), 289–316.

Shepard, R. N., C. L. Hovland, and H. M. Jenkins (1961). Learning and memorization of classifications. *Psychological Monographs 75*(13, Whole No. 517).

Silverman, D. (1999, April). On allophonic relations: phonetic similarity or functional identity. Handout, GLOW99 (Generative Linguistics in the Old World) Workshop on Phonetics in Phonology.

Silverman, D. (2006). *A critical introduction to phonology: of sound, mind, and body.* Critical Introductions to Linguistics. Continuum Books.

Skoruppa, K., A. Lambrechts, and S. Peperkamp (2009). The role of phonetic distance in the acquisition of phonological alternations. In S. Lima, K. Mullin, and B. Smith (Eds.), *Proceedings of the 39th Meeting of the North-East Linguistics Society (NELS)*, Amherst, Massachusetts. Graduate Linguistics Students Association.

Skoruppa, K. and S. Peperkamp (2011). Adaptation to novel accents: feature-based learning of context-sensitive phonological regularities. *Cognitive Science 35*, 348–366.

Smith, J. D., J. P. Minda, and D. A. Washburn (2004). Category learning in rhesus monkeys: a study of the Shepard, Hovland, and Jenkins (1961) tasks. *Journal of Experimental*

Psychology: General 133(3), 398–404.

Sproat, R. and O. Fujimura (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics 21*, 291–311.

Walsh Dickey, L. (1997). *The phonology of liquids*. Ph. D. thesis, University of Massachusetts, Amherst, Massachusetts.

Warker, J. A. and G. S. Dell (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition 32*, 387–398.

Warker, J. A., G. S. Dell, C. A. Whalen, and S. Gereg (2008). Limits on learning phonotactic constraints from recent production experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition 34*(5), 1209–1295.

Wilson, C. (2003). Experimental investigation of phonological naturalness. In G. Garding and M. Tsujimura (Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics (WCCFL 22)*, Somerville, pp. 533–546. Cascadilla Press.

Yu, A. C. L. (2010). Tonal effects on perceived vowel duration. In C. Fougeron, B. Kühnert, M. D'Imperio, and N. Vallée (Eds.), *Papers in Laboratory Phonology 10: phonology and phonetics*, pp. 151–168. Mouton De Gruyter.

Yu, A. C. L. (2012). On measuring phonetic precursor robustness: a response to Moreton. *Phonology 28*(3), 491–518.