Glossa a journal of general linguistics

Avcu, Enes and Arild Hestvik. 2020. Unlearnable phonotactics. *Glossa: a journal of general linguistics* 5(1): 56. 1–22. DOI: https://doi.org/10.5334/gjgl.892

RESEARCH

Unlearnable phonotactics

Enes Avcu[1] and Arild Hestvik[2]

[1] Department of Neurology, The Massachusetts General Hospital, Harvard Medical School, Boston, MA, US

[2] University of Delaware, Newark, DE, US

Corresponding author: Enes Avcu (eavcu@mgh.harvard.edu)

The Subregular Hypothesis (Heinz 2010) states that only patterns with specific subregular computational properties are phonologically learnable. Lai (2015) provided the initial laboratory support for this hypothesis. The current study aimed to replicate and extend the earlier findings by using a different experimental paradigm (oddball task) and a different measure of learning (sensitivity index, *d'*). Specifically, we compared the learnability of two phonotactic patterns that differ computationally and typologically: a simple rule ("First-Last Assimilation") that requires agreement between the first and last segment of a word (predicted to be unlearnable), and a harmony rule ("Sibilant Harmony") that requires the agreement of features throughout the word (predicted to be learnable). The First-Last Assimilation rule was tested under two experimental conditions: one where the training data were also consistent with the Sibilant Harmony rule, and one where the training data were only consistent with the First-Last rule. As in Lai (2015), we found that participants were significantly more sensitive to violations of the Sibilant Harmony (SH) rule than to the First-Last Assimilation (FL) rules. However, unlike Lai (2015), we also found that participants showed some residual sensitivity to the First-Last rule, but that sensitivity interacted with rule type so that participants were significantly more sensitive to SH rule violations. We conclude that participants in Artificial Grammar Learning experiments exhibit evidence of Universal Grammar constraining their learning, but patterns predicted to be unlearnable as a linguistic system can still be learned to some degree, due to non-linguistic learning mechanisms.

Keywords: Phonotactics; Learnability; Computational Complexity; Subregular Hypothesis; Domain Specificity

1 Introduction

The perennial question in phonology is why some patterns are observed in languages and others not. Moreton (2008) addresses this question by discussing two proposals: the first is analytic bias — the presence of cognitive filters that help the learning of some patterns while suppressing others (Wilson 2003). Universal Grammar can be thought of as an example of analytic bias in which innate mechanisms facilitate the learning of a certain set of structural rules (Moreton, 2008). The other proposal is the channel bias — the presence of phonetically systematic errors in transmission between the speaker and learner (Ohala 1993; Hale & Reiss 2000). The perceptual similarity between sounds has been argued to be one of the sources of channel bias (Ohala 1993). In addition to these proposals, phonologists have debated to what extent learnability can explain why some sound patterns are attested while others are not; and have explored how factors such as complexity and naturalness affect the learnability of a sound pattern (Moreton 2008; Heinz 2010; Heinz & Idsardi 2013). Heinz (2010) suggests that the absence of some patterns in phonology is due to learnability constraints which can be described in terms of computational complexity. Many patterns that are unlearnable are outside the range of certain complexity patterns, and only patterns within this subclass are learnable.

These claims can be tested in the laboratory by comparing the learnability of two patterns that are similar on the surface, but different both typologically and computationally. Sibilant Harmony patterns are attested empirically and therefore necessarily fall in the class of languages that should be learnable. On the other hand, what we call First-Last Assimilation patterns are unattested and fall outside specific computational complexity classes. This learnability difference has been observed in a previous artificial language learning experiment (Lai 2015). The current study aimed to replicate these findings, but expand empirical coverage by using a different experimental paradigm.

The Sibilant Harmony (SH) rule requires all sibilants of a word to agree in the [*anterior*] feature and is an attested pattern in Chumash (Applegate 1972) and Navajo (Sapir & Hoijer 1967). The First-Last Assimilation (FL) rule requires only the first and last sibilants of a word to agree, but this pattern is not attested in any human language. Heinz and Idsardi (2013) argue that the patterns present or absent in phonology cannot be explained by the general psychological mechanisms such as working memory or perception. For example, consider the fact that the first and last sound of a word are relatively salient (Endress & Mehler 2010). From a saliency perspective, it would seem plausible that language could have a harmony rule that requires the first and last sounds of a word to agree (FL). For example, sibilants in these positions should be perceptually more salient than sibilants targeted by a sibilant harmony rule. However, this assimilation pattern is nevertheless unattested among the world's languages. More interestingly, the FL pattern does not belong to the specific subregular classes of the subregular hierarchy that include observed phonotactic patterns. Heinz and Idsardi (2013) proposed that the absence of some patterns in the phonology of the world's languages is due to the computational complexity of those patterns making them unlearnable.

The artificial grammar learning (AGL) paradigm offers a way to test this hypothesis in laboratory settings. Lai (2015) found empirical evidence for the Subregular Hypothesis by using AGL. AGL consists of a training phase, followed by a testing phase. In the training phase, participants are exposed to an artificial grammar constructed by the researcher, and the test phase measures whether they learned the pattern or the rule system of the artificial grammar. Lai (2015) compared the learnability of Sibilant Harmony (SH) vs. First-Last Assimilation (FL) and found that the FL pattern was not learned, but the SH pattern, which belongs to a specific subregular region of the complexity hierarchy, was learned. The current study aimed to add new experimental evidence for the learnability of the SH pattern over the FL pattern, as hypothesized by Heinz (2010) and confirmed by Lai (2015). Section 1.1 below presents the formal and computational background of the hypothesis, Section 1.2 details the comparison between the Sibilant Harmony and First-Last Assimilation patterns, and Section 1.3 reviews previous work. In Section 1.4, we layout our motivation and contribution with an argument for the importance of replicating Lai (2015).

In our study, we used the same training phase as in Lai (2015), but a different testing method. Specifically, we used the *oddball task* (presentation of infrequent ungrammatical stimuli among frequent grammatical stimuli (the oddball task was chosen as it was part of another study using EEG and ERPs, not reported here)). We also used Signal Detection Theory (SDT) (Green & Swets 1966; Macmillan & Creelman 2004), namely the sensitivity index *(d')* to measure learning. SDT is a data analysis tool that categorizes stimuli into signal and noise, and the sensitivity index measures whether and how good the participants are at detecting signals given the background noise and observer uncertainty. Learning was operationalized as high sensitivity to ungrammatical forms in the testing phase. The experimental details are explained in Section 2.

To preview our results, presented in Section 3, we replicated the earlier findings that an attested and computationally learnable pattern (SH) is inside the hypothesis space of humans' phonological pattern detectors. What is new in the current study is that we also found a residual level of sensitivity to the unattested and predicted-to-be-unlearnable FL patterns. We suggest that this reveals traces of a psychological domain-general learning mechanism, existing alongside with innate, domain-specific language learning mechanisms. As such, our findings agree with Musso et al. (2003), who tested learning of both natural and non-existing syntactic patterns, and found that both attested and unattested syntactic rules (where the latter violated principles of Universal Grammar) were learned when behavioral measures were used. However, only the attested rules showed activation in language-related brain areas (using fMRI). Our findings of weak learning of unattested FL rules extend this type of observation to the domain of phonology. Implications of these results are discussed in Section 4, and Section 5 concludes the paper.

1.1 Background: the Subregular Hypothesis

The Chomsky Hierarchy (Chomsky 1956) divides all logically possible patterns into nested regions of complexity. Each of these regions has mathematical definitions that enable any machine or algorithm to generate the strings comprising the pattern (Harrison 1978; Hopcroft, Motwani & Ullman 2006). Also, each region specifically distinguishes abstract, structural properties of grammars — i.e., a machine with finitely many internal states can only recognize patterns belonging to the regular region. As can be seen in Figure 1, different regions contain different linguistic generalizations which are modeled as stringsets. The regions from context-sensitive to context-free contain syntactic phenomena like relative clause copying in Yoruba (Kobele 2006), cross-serial dependencies in Swiss German (Shieber 1985), and nested dependencies in English (Chomsky 1957).

Phonological patterns reside in the regular region (Johnson 1972; Kaplan & Kay 1994). The regular region is the smallest subset of this hierarchy, and it contains finite stringsets. Heinz (2018) notes that "the primary result in computational phonology to date is that the transformations from underlying to surface forms […] are in fact regular" (Heinz

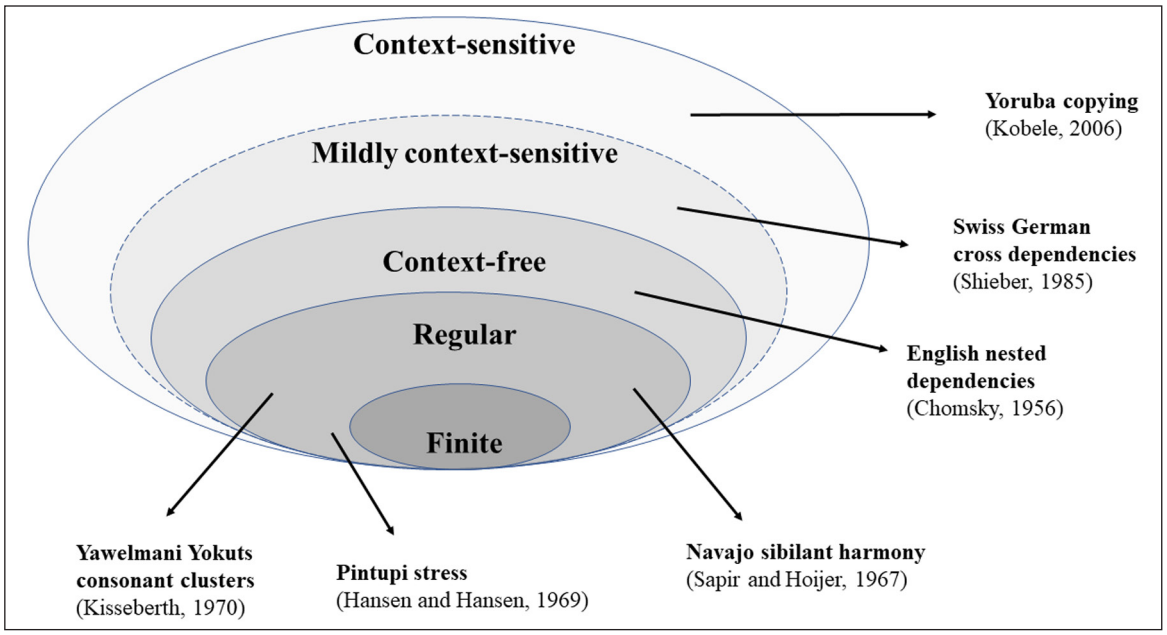

Figure 1: The Chomsky Hierarchy. Various features of natural language occupy different regions of the hierarchy. Figure reproduced from Figure 1 in Heinz (2010: 634) with permission.

2018: 139). For example, phonological phenomena like the constraint on adjacent consonant clusters in Yawelmani Yokuts (Kisseberth 1970), Pintupi stress patterns (Hansen & Hansen 1969), and Navajo sibilant harmony (Sapir & Hoijer 1967) can be modeled by finite grammars. Although all phonological generalizations are regular, not all regular patterns are "phonological" — meaning that phonological patterns are part of a specific subset of regular formal languages. For example, FL is a regular logically possible pattern but it is not phonological (Heinz & Idsardi 2013).

The Subregular Hierarchy, a subcategorization of the regular region, divides regular patterns into classes of different complexity (McNaughton & Papert 1971; Rogers et al. 2010; 2013; Rogers & Pullum 2011; Heinz & Rogers 2013). Heinz (2010) showed that phonotactic patterns in natural languages inhabit proper subsets within the regular region.[1] These subsets are the Strictly-Local, Strictly-Piecewise, and Non-Counting regions (or Locally Testable with Order) (McNaughton & Papert 1971; Heinz 2010; Rogers et al. 2010; Rogers & Pullum 2011; Heinz & Rogers 2013).

A strictly *k*-Local (SL*k*) pattern is one in which the well-formedness of a string is determined by whether its contiguous substrings of length *k* are well-formed (where *k* is the window of segments over which the restriction is regulated). Strictly local languages only make distinctions based on contiguous substrings up to some length *k* (called *k*-factors). Strictly *k*-local grammars can be thought of as n-gram models in computational theory. The class of strictly *k*-local languages is known to represent the phonological patterns of spreading and correspondence restrictions in natural languages (Heinz 2010). Co-occurrence restrictions in phonology belong to this class — a rule like **ab* can be described as a strictly 2-local pattern which restricts the co-occurrence of *a* immediately followed by *b* (note that *a* and *b* are adjacent, thus the dependency is local).

A strictly *k*-Piecewise (SP*k*) pattern, on the other hand, is one where the well-formedness of a string is determined by its subsequences (non-adjacent strings) of length *k*. If the set of subsequences in the string in question is a subset of the set of subsequences allowed by the grammar, the string is well-formed; otherwise, it is not. Thus, subsequences are not necessarily adjacent; the patterns they describe contain long-distance dependencies. The class of strictly *k*-piecewise languages is known to represent the phonological patterns of symmetric and asymmetric long-distance patterns like consonantal harmony. A rule like **a…b* can be described as a strictly 2-piecewise pattern which restricts the occurrence of *a* followed by *b*. (Note that *a* and *b* are non-adjacent, therefore the dependency is not local.) A linguistic example would be the sibilant harmony rule in Navajo, in which a word may not contain two sibilants with differing anteriority features (Sapir & Hoijer 1967). Thus, a word like [ʃi-d͡ʒaa] 'a mass lies' is following the rule, while an artificial word like [si-d͡ʒaa] is violating it. This sibilant harmony rule can be modelled as a strictly 2-piecewise rule because the grammaticality of the word can be checked by observing the 2-factors in the word. Since each 2-factor {ʃ…i, ʃ…d͡ʒ, ʃ…a, i…d͡ʒ, i…a, d͡ʒ…a, a…a} is following the rule, this word is well-formed. For the strictly piecewise class, the order of segments is important, but not the distance between them. Most attested cases of consonant harmony can be characterized as strictly 2-piecewise (Heinz 2010).

In the regular region, apart from strictly local and strictly piecewise classes, there are also other regular patterns which are neither strictly local nor piecewise. These patterns can be subsumed under the Non-Counting patterns, also called Star-Free and Locally Testable

[1] Recent work has provided evidence that phonological rules such as epenthesis, metathesis, and deletion can be modeled with subregular relations (Chandlee 2014). In addition, the markedness constraints in Optimality Theory (Prince & Smolensky 1993; 2004) (Hayes & Wilson 2008), long-distance phonotactic constraints (Heinz 2010), most of the stress patterns in the world's languages (Edlefsen et al. 2008), and phonological tone patterns (Jardine 2016) can be described by specific subregular constraints.

with Order. A pattern is Non-Counting if there is a number n such that for all strings u, v, w, if uv^nw occurs in L, then $uv^{n+1}w$ occurs in L as well (McNaughton & Papert 1971).

To summarize, at the top of the subregular hierarchy is the regular region, and at the bottom is the finite region. Under the regular region, there is the Non-Counting region which is dominating the Locally Threshold Testable, Locally Testable and Piecewise Testable regions. These intermediate regions are between the strictly local, strictly piecewise and non-counting regions. According to Heinz (2018), the First-Last Assimilation rule specifically belongs to the Locally Testable class. Zalcstein (1972) defines this class as a language that is expressible as a boolean combination of strictly local languages. Figure 2 presents a schematized representation of the subregular classes.

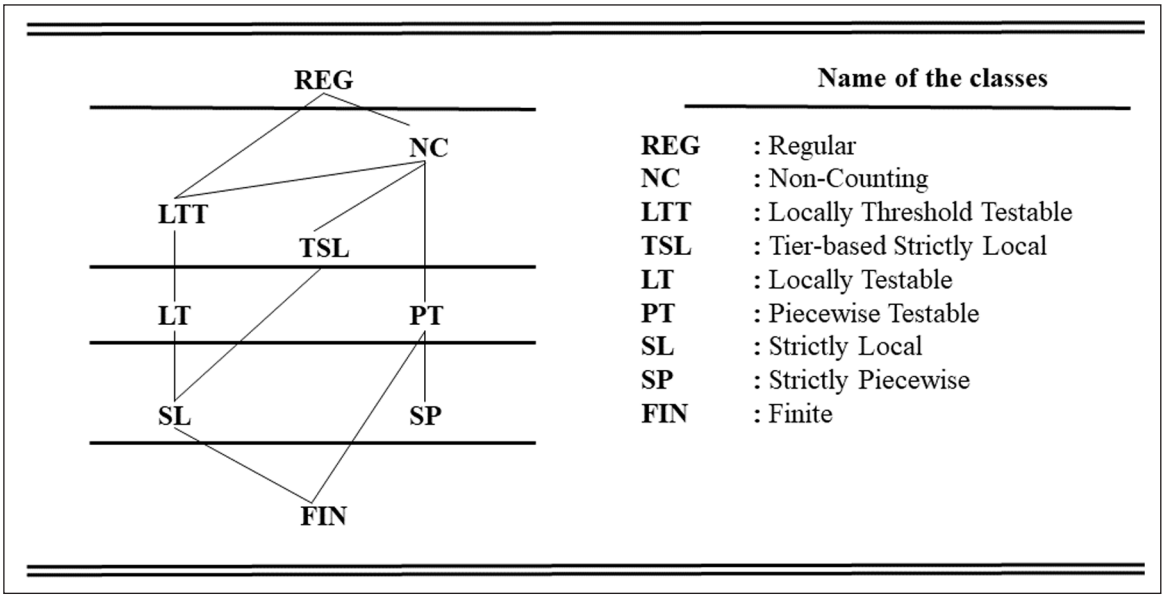

Figure 2: Subregular Boundaries. Strictly-Local, Strictly-Piecewise, and Non-Counting classes are all in the regular region. Figure reproduced from Figure 2 in Heinz (2018: 14) with permission.

In contrast to the Non-Counting patterns, the Strictly Local and Strictly Piecewise classes include almost all-natural language phonotactic patterns (Heinz 2010); that is, no language has a phonotactic pattern like First-Last Assimilation. In this respect, Heinz's (2010) Subregular Hypothesis is supported by the typology of phonotactic patterns and suggests that humans' phonological pattern detectors are limited to detecting grammars that are Strictly-Local or Strictly-Piecewise. If this is the case, then the absence of patterns from natural languages such as the First-Last Assimilation can be explained; namely, the regularities present in the patterns of the First-Last Assimilation cannot be extracted by humans' phonological learning mechanism. In other words, patterns with specific subregular computational properties are privileged with respect to learnability.

1.2 The comparison between the Sibilant Harmony and First-Last Assimilation patterns

From the perspective of formal logic, Sibilant Harmony (SH) can easily be defined as the conjunction of negative literals. It can also be defined as "**s…ʃ* and **ʃ…s*" markedness constraints in Optimality Theory. On the other hand, markedness constraints for First-Last Assimilation (FL), "**#s…ʃ#* and **#ʃ…s#*" must include the position symbols, because of the effect of position on the grammaticality of the word. As for the computational complexity of these patterns, SH belongs to the Strictly Piecewise class, while FL belongs to the Locally Testable class (which is subsumed under the Non-Counting patterns). The difference between these two patterns becomes apparent when a word has at least three

sibilants. In this case, when the medial sibilant disagrees with the other two sibilants, the word is grammatical according to FL but violates the SH pattern. (It is also important to note that SH is a proper subset of FL — a word in the SH language is also a part of the FL language.)

Although the FL pattern is not attested in human languages, there are cases similar to FL discussed in the phonology literature. Finley (2009) discusses a possible FL agreement rule as a morpheme realization rule and tested the hypothesis that such patterns are only learnable as morphological alternations (Finley 2012a). A very similar attested pattern is the vowel harmony pattern in C'Lela, a Niger-Congo language spoken in Nigeria (Archangeli & Pulleyblank 2007; Dettweiler 2000; Pulleyblank 2002). In C'Lela, the vowels in the root and the final suffix agree in height, ignoring the non-final suffixes which have become transparent after a process called suffix stacking. Thus, the trigger of the vowel harmony is the vowel in the root, and the target is the final suffix. However, the interpretation of an edge-sensitive vowel harmony pattern in C'Lela does not make it an FL pattern, because both the motivation (trigger factor) and the target of the assimilation process depend on the position (Lai 2015).

Another FL-like pattern was reported in Endress & Mehler (2010) where participants learned a phonotactic constraint expressed by the following rule: The consonants C_1 and C_2 in words of the form C_1VCCVC_2 must come from two distinct sets: {k, t, f} (Set 1) and {s, ʃ, p} (Set 2). Endress & Mehler (2010) found that participants were able to learn this pattern. However, the pattern they tested was not FL; it was a Strictly Local pattern with #k, #t, #f permissible and #s, #ʃ, #p forbidden, or vice versa. The study showed that within the Strictly Local class, learning a constraint like "#s is forbidden" is easier to learn than a constraint like "*sf* is forbidden". As Endress & Mehler (2010) themselves pointed out, "…participants did not learn any relation among consonants at all; rather, they just had to remember the positions in which each consonant could occur" (Endress & Mehler 2010: 240). However, FL requires learning a position-based *relation* among consonants.[2]

A final pattern that is similar to FL was discussed by Koo & Callahan (2012), where a long-distance dependency pattern can be interpreted as position-bound. Participants in this study were able to learn a phonotactic constraint where the consonants C_1 and C_3 in words of the form $C_1VC_2VC_3V$ had occurrence restrictions: (i) when C_1 is [s], C_3 cannot be [l] and (ii) when C_1 is [g], C_3 cannot be [m]. The fact that this pattern describes an arbitrary relation between the consonants rather than an assimilation process, is what differentiates it from the FL pattern (Lai, 2015). Therefore, even though the attested vowel harmony pattern in C'Lela, and the artificial patterns in Endress & Mehler (2010) and Koo & Callahan (2012) seem similar to the FL pattern, neither Endress & Mehler (2010) nor Koo & Callahan (2012) tested a rule which has the computational properties of FL and the naturalness of an attested pattern (harmony).

1.3 Behavioral evidence for the Subregular Hypothesis

Many studies have used the AGL paradigm to test the learnability of language patterns (Marcus, Vijayan, Rao, & Vishton 1999; Öttl, Jäger & Kaup 2015) and specifically phonology (e.g., Peperkamp, Le Calvez, Nadal & Dupoux 2006; Lai 2015; Finley 2017). Remarkably, it has been shown that after a very brief training session, both seven-month-old infants (Marcus et al. 1999) and sixteen-month-old infants (Chambers, Onishi & Fisher 2003), as well as adults are able to learn the grammar and the phonotactics of an artificial language (Onishi, Chambers & Fisher 2002). While many types of phonotactic patterns present dependencies between adjacent segments (strictly local in terms of subregular

[2] We thank Jeffrey Heinz for pointing this out.

complexity), many phonotactic patterns result from interactions between non-adjacent segments with intervening elements (consonant or vowel harmony patterns that are strictly piecewise in terms of subregular complexity). The learnability of strictly local patterns has been well studied in laboratory settings (Aslin, Saffran & Newport 1998; Dell et al. 2000; Chambers, Onishi & Fisher 2003; Goldrick 2004; Onnis et al. 2005; see Cristia (2018) for a contrasting view). In most of these studies, it has been observed that, by employing statistical learning methods, both infants and adults use phonotactic regularities to segment words from a continuous stream of an artificial language.

Although the learnability of strictly piecewise patterns poses different challenges to the learner due to their inherent complexity, it has been shown that they are learnable in laboratory settings (Pycha, Nowak, Shin & Shosted 2003; Wilson 2003; Newport & Aslin 2004; Onnis et al. 2005; Finley & Badecker 2009a; b; Finley 2011; 2012b; Koo & Callahan 2012). Pycha et al. (2003) compared the learnability of a vowel harmony rule to a vowel disharmony rule — the latter of which is not frequently found in human languages. The results showed that participants learned both the harmony and disharmony patterns, and there was no significant difference between the two. However, note that the explicit feedback given to participants during the learning phase of both patterns might have induced the participants to use a different learning strategy than the one used in natural settings. Pycha et al. (2003) also tested whether participants could learn an arbitrary vowel dependency rule and showed that the learnability of the arbitrary rule was worse than the harmony and disharmony rules. Similarly, Wilson (2003) found that when participants were tested on assimilation and dissimilation processes compared to a random process, they were better at learning the former. Both these studies show that when participants are tested on unnatural patterns, they learn the natural patterns better.

In a statistical learning experiment, Newport & Aslin (2004) compared the ability to segment a word from a continuous speech stream using transitional probabilities. Results demonstrated that participants successfully segmented words when the dependency was between two non-adjacent segments, but not between syllables. Finley (2011) reported that when participants were tested on the learnability of a sibilant harmony pattern in which long-distance dependencies with different distances were controlled by the number of intervening segments, their learning was locally biased. This means that shorter-distance patterns were preferred over long-distance patterns. Lai (2012) discussed this as evidence for the subregular complexity hypothesis, in that the usage of a strictly local learner is prioritized over a strictly piecewise learner. In a follow-up study, Finley (2012) showed that when participants were trained on long-distance patterns with varying complexity (again depending on the number of intervening elements between the two segments), they were able to generalize beyond the training set and learn the long-distance dependencies in an unbounded way. In another study, Finley (2015) tested the learnability of a long-distance vowel harmony pattern. The results showed that when the pattern included a transparent vowel that makes the dependency more complex, participants required extra training to learn the harmony pattern. Koo & Callahan (2012) also reported learnability results from long-distance harmony patterns. In their study, the dependency between [s] and [l], and [g] and [m] were tested in trisyllabic words with the structure of *C.V.C.V.C.V.* It was reported that participants preferred novel legal words over novel illegal words. The results suggested that the dependency between the two sounds is learned by ignoring the actual distances between segments. In other words, both strictly local and strictly piecewise patterns have been shown to be learnable in laboratory settings.

But what about the patterns that are subregular but neither strictly local or strictly piecewise? Lai (2015) examined this question by comparing the learnability of two

long-distance harmony patterns with an artificial grammar learning paradigm and tested whether SH or FL can be learned by adult participants in a laboratory setting. Three experimental groups were tested (SH, FL, and a control group with no training phase). The two test groups underwent two phases: a training phase and a testing phase. The SH group was trained by listening to words that conformed to an SH grammar, and the FL group was trained by listening to words that conformed to an FL grammar. The control group received no training. In the test, a two-alternative forced-choice (2AFC) task was used. Participants had to judge whether the first word or the second word of a pair were more likely to belong to the artificial language they had previously been exposed to. Participants in the control condition (which were not given a training phase) were simply asked to judge whether they thought the first or the second word of each pair was a better candidate for a possible word. All participants were given the same test stimuli.

The results of Lai's study showed that the experimental group that was trained on the SH pattern preferred the words following the SH rule over the ones that violated it. Thus, the SH rule was learned by the participants. On the other hand, the FL participants did not show any preference for the FL rule — they did not perform significantly better than the control group. This suggests that FL grammars are indeed unlearnable. Interestingly, Lai also observed that the FL group showed a preference for stimuli that conformed to the SH pattern, i.e. a bias towards SH-conforming words. Lai speculated that they may have learned the SH pattern from the FL stimuli. A possible explanation for this is that anything that violates FL also violates SH, and anything that conforms to SH also conforms to FL, cf. Figure 3.

Therefore, given the same experimental setting and the same amount of training, the FL group appeared to learn SH grammar when exposed to FL stimuli. To address this potential SH bias, Lai designed a follow-up experiment in which the FL participants were trained with stimuli that conformed *only* to the FL pattern. Thus, the [s.s.s] and [ʃ.ʃ.ʃ] type of words was excluded from the training set, leaving only the [s.ʃ.s] and [ʃ.s.ʃ] type of words. The results of this follow-up experiment showed that when participants were trained with these "intensive" FL (henceforth "IFL") stimuli, they preferred the stimuli that conformed only to the IFL pattern. In other words, after removing the ambiguous stimuli, the IFL group internalized a sibilant disharmony rule which requires each neighboring sibilant to be disharmonic. Lai (2015) concluded that the sum of the experiments indicated that SH, not FL was learned. These results were consistent with the hypothesis that the phonological learner is restricted by sub-regular constraints to learn SH, but not FL.

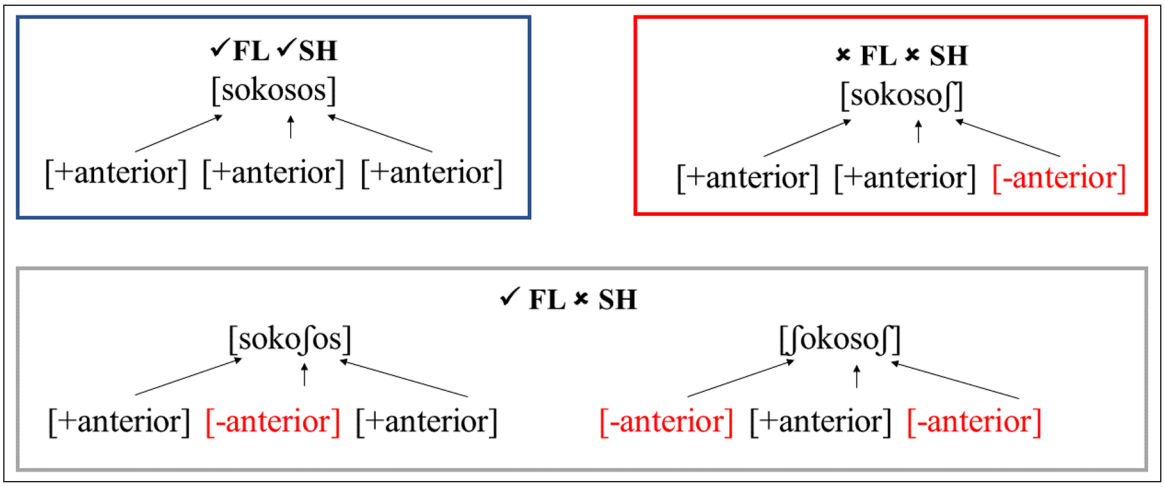

Figure 3: Comparison of SH and FL stimuli.

1.4 The current study

To repeat, the current study aimed to replicate Lai's learnability results, but with a different test design: an oddball paradigm; and with a different measure: the sensitivity index (d') as defined by Signal Detection Theory (SDT). In our study, an ungrammatical word form (either according to SH or FL) is conceived of as the "signal" that the listener is tasked to detect. The size of the participant's d', their sensitivity, measures how sensitive he/she is to ungrammaticality. If a pattern is not learned, then the participant's sensitivity to ungrammaticality should be zero. On the other hand, different degrees of positive sensitivity can be taken to reflect how stable the grammatical knowledge is. Another important aspect of using SDT is that it factors out the participant's response bias (c) from their sensitivity. In our paradigm, the probability of encountering ungrammatical strings is lower than the probability of encountering grammatical forms. This leads to a bias towards expecting grammatical forms; SDT allows us to factor out this bias from our grammatical knowledge measure.

We assume that once a learner has extracted a rule from a set of training data, the psychological processing system implements the rule and starts to generate predictions during real-time phonological parsing: New and subsequent input should conform to the rule. During parsing of a word, an error signal is generated in the brain if the rule-based predictions about the phoneme sequence in the word are not met. This signal is informing the participant's judgment and eventually is translated into a behavioral response. If a participant fails to learn a rule (e.g. the language-impossible FL rule), this should be reflected in a lack of predictions at the phonological processing level, and participants will not detect the signal — i.e., they will have low sensitivity to the presence of ungrammatical word forms. Sensitivity is also a less theory-laden concept compared to grammaticality judgments: grammaticality judgments imply that the participant has a concept of well-formedness, which is not necessarily clear to naïve participants. Sensitivity, on the other hand, merely asks the participants to judge whether a given word form was perceived as different or "not belonging to the language" they had just learned—a perceptual measure.

2 Method

2.1 Participants

A total of 72 University of Delaware students were recruited as participants, divided into three groups with 24 participants in each group. Each participant received course credit for participation. 66 of the 72 participants were females and 6 were males (the imbalance arises from the overrepresentation of women in our sampling population). Six participants were left-handed. The mean age was 22 (SD = 4.32, range = 18 to 31). None of the participants reported a history of hearing loss or speech/language impairments, and all reported having English as their first and only language. Informed consent for this study was obtained in compliance with the Human Subjects Review Board at the University of Delaware (IRB 811097-1).

2.2 Stimuli

The study consisted of three experimental conditions. The first tested the learnability of the Sibilant Harmony (SH) rule, and the second tested the unattested First-Last Assimilation (FL) rule. The third condition tested the learnability of FL under an "intensive" condition; the Intensive First-Last Assimilation (IFL) rule, which is like the FL condition except for training items consistent with SH is omitted.[3] No control group was used in the current

[3] Intensive FL specifically belongs to the Tier-Based Strictly Local (TSL) class which is a specific generalization of Strictly Local class and defined with a phonological tier. Similar to the strictly local class, TSL class can be defined with conjunctions of negative literals after non-tier elements are ignored. For example, in a

study because Lai (2015) already demonstrated that a group with no training or random training will display zero sensitivity.

We used the exact same stimulus recordings as in Lai (2015). All the training and test stimuli had three syllables in the form of "*CV.CV.CVC*". The consonants in the inventory of the language were only [k,s,ʃ], and the vowels were [a,ɛ,ɔ,i,u]. Half of the training stimuli had a [k] as the second consonant and the other half had a [k] as the third consonant. Therefore, the first and last consonants were always sibilants. In the testing phase, disharmonic words for each condition were in four different forms: For the SH condition, the disharmonic sibilant was either [s] or [ʃ] and the position of this sibilant was either the second or the third sibilant. For the FL conditions, the disharmonic sibilant was [s] or [ʃ] and this sibilant could be different from the second sibilant or the same. All of the words which had a disharmonic sibilant at the end ([s.s.ʃ] or [ʃ.ʃ.s]) had a [k] as the second consonant. Half of the words which had a disharmonic sibilant in the middle ([s.ʃ.s] or [ʃ.s.ʃ]) had a [k] as the second consonant and the other half as the third consonant. The mean duration of stimuli was 1013ms; the longest stimulus was 1251ms and the shortest was 884ms. Table 1 summarizes the types of training and test stimuli.

2.3 Apparatus and procedure

The experiment was programmed with E-Prime Professional software v. 2.0.10.356, running on a Dell desktop PC. The experiment was conducted inside a single-walled shielded sound booth in the Experimental Psycholinguistics Lab at the University of Delaware. The presentation of sound stimuli was executed with two free field speakers with dual-mono presentation, placed in front of the participants at comfortable listening volume (loudspeakers placed at 45° angles approximately 1 m in front of the participant). Visual input (e.g., instructions) was delivered through an LCD screen placed on a table in front of the participant. The PST Serial Response box was used for recording behavioral responses.

The procedure consisted of two phases: a training phase and a testing phase. During the training phase, participants listened to grammatical words and were instructed to repeat each word orally once they heard it. The training session contained 200 tokens (40 words

Table 1: Stimulus patterns across three experimental conditions: SH, FL, and IFL. A checkmark indicates that a particular type of stimulus is included in the stimuli list.

	Experimental Conditions					
	SH		FL		IFL	
Sibilant Tier	Harmonic	Disharmonic	Harmonic	Disharmonic	Harmonic	Disharmonic
[s.s.s]	✓		✓			
[ʃ.ʃ.ʃ]	✓		✓			
[ʃ.s.ʃ]		✓	✓		✓	
[s.ʃ.s]		✓	✓		✓	
[ʃ.ʃ.s]		✓		✓		✓
[s.s.ʃ]		✓		✓		✓
[ʃ.s.s]				✓		✓
[s.ʃ.ʃ]				✓		✓

hypothetical word like *sakasas*, when the vowels are ignored, the sibilant tier [s.s.s] holds a local relation which is a limited kind of long-distance behavior, as noted by Heinz (2018). See Heinz et al. (2011) for a more formal definition of TSL languages and proofs for several computational properties of the TSL class.

× 5 repetitions) and the duration was approximately 15 minutes. The training phase was an exact replication of Lai (2015). The training was followed by a testing phase. Stimuli were presented in an oddball paradigm, where an ungrammatical stimulus appears infrequently (18% of the time) among occurrences of grammatical stimuli (82% of the time). Participants were presented with the words in a continuous stream and were asked to "press the button when you think the word you heard does not belong to the language you had just learned during training." The testing phase presented a total of 528 trials: 432 grammatical words (72 tokens × 6 repetitions) and 96 ungrammatical words (12 × 4 tokens × 2 repetitions). The test phase was divided into two blocks, each of which had the same total number of trials. A random number of grammatical words (between 3 and 7) occurred between each ungrammatical word. Stimuli were delivered in two blocks, and the 264 trials in each block consisted of 48 ungrammatical (18%) and 216 grammatical (82%). The total duration for both training and testing was about 50 minutes.

The task for the participant was to detect ungrammatical stimuli by pressing a response box button to indicate when this occurred. Participants only pressed the button for ungrammatical words. No explicit feedback was given to participants during the test phase because this would provide additional learning cues during testing. The testing phase was thus completely different from Lai (2015) where pairs of words were presented, and accuracy was collected.

2.4 Data recording

Due to the nature of this specific task, which is detecting the signal (ungrammatical word) against the background noise or non-signals (grammatical words), Signal Detection Theory (SDT) (Macmillan & Creelman 2004), was used to analyze the results. SDT is widely used in psychology (e.g., psychophysics, perception, memory or statistical decision), it can be applied to any type of discrimination task where two possible stimuli must be discriminated (Stanislaw & Todorov 1999), and is widely used in speech perception experiments (Keating 2004). However, the use of SDT in artificial language learning experiments and grammaticality judgments is novel to the current study. In SDT terms, subject responses can be classified into four classes: hits, false alarms, misses, and correct rejections. In the current experiment, the signal detection scenario is described in the Table 2.

To compute the sensitivity index, d', only hits and false alarms are needed, as missed and correct rejections are the complement and therefore contain the same (if inverse) information. In the test phase, button presses made by participants to ungrammatical stimuli were recorded. When the signal (ungrammatical words) was present and the participant detected it and reported hearing it, it was counted as a hit. The proportion of hits was calculated as $P(H) = Nhits/Nsignals$. with N being the number of times that the event was observed. When the signal was absent, but the participant still thought they observed something and reported it (i.e., when a grammatical word was presented, but the participant reported it as ungrammatical), it was counted as a false alarm. The proportion of false alarms was calculated as $P(FA) = Nfalsealarms/Nnosignals$. The sensitivity index is then calculated as $d' = Z(P(H)) - Z(P(FA))$, where $P(H)$ is the proportion

Table 2: Signal detection scenario in the current experiment.

		Stimulus type	
	Signal	**Ungrammatical (the "signal")**	**Grammatical ("no signal")**
Subject's decision	"I noticed a signal"	Hit	False alarm
	"There is no signal"	Miss	Correct Rejection

of hits, *P(FA)* is the proportion of false alarms, and *Z* is the z-score for those proportions or probabilities.[4]

The bias measure (*c*) represents participants' positive or negative bias towards making a "signal" decision and is derived from the hit and false alarm rates, calculated as $c = (Z(Phits) + Z(Pfalsealarms))/2$. The bias measure reflects the balance between false alarms and misses: when the false alarm and miss rates are equal, *c* equals zero; if false alarm rates are higher than the misses, there is a positive bias and participants are "aggressive"; when there are more misses than false alarms, the bias is negative, i.e. participants are "conservative." This illustrates the advantage of using SDT: Participants may be biased towards thinking that most words are grammatical, and this bias can come from multiple sources: as a consequence of the low probability of ungrammatical words, as well as an expectation that language examples should be grammatical, which is natural in language acquisition; learners expect other people to speak grammatically. Using SDT allows us to factor out this bias from the participants' sensitivity.

The d' and *c* parameters differentiate sensory factors from decision factors (DeCarlo 1998). When participants cannot discriminate the signal from the noise at all, hits would be equal to false alarms, which gives $d' = 0$. Results higher than 0 show that sensitivity is better than chance level. Thus, in the context of our study, a positive d' means the rule is learned (in the sense that performance is better than guessing). Furthermore, the higher the d', the more confident the learner is about the rule or the better they are at detecting violations of the rule. This is how learning is defined using SDT within our experimental context.

2.5 Analysis

The mean sensitivity (d') and bias (c) for each of the three groups were computed and used as dependent measures in statistical tests. We conducted both a non-parametric version of the one-sample t-test, the Wilcoxon Signed-Rank test, along with inferential statistics with logit transformed hit- and false alarm rates. d' scores were not tested directly by ANOVA, because—as pointed out by a reviewer—the normality assumption does not appear hold for means with d' values close to zero. The reason is that sensitivity is conceptually bounded at the lower end by 0—zero sensitivity means that a participant would have to resort to guessing whether a signal is present or not (akin to a blind person making decisions about whether a flash of light was presented or not). Therefore, d' scores are not expected to be distributed symmetrically around its mean when that mean is close to zero.[5] Given this lower bound of zero, the question arises whether mean d' scores close to zero can be assumed to be sampled from a normally distributed theoretical sampling distribution of means, i.e. where one tail of the sample means would cross over the 0 point and be negative. However, since d' is assumed to only be meaningful from zero and up, this situation would violate normality assumptions of ANOVA.

For this reason, we conducted a statistical analysis of hit rates and false alarm rates by converting these probabilities to their corresponding log-odds (the "logit"), as is commonly done with proportions as dependent measures. As shown by DeCarlo (1998), signal detection models can be formulated as a subclass of generalized linear models (GLMs),

[4] When the probability of hit and false alarm becomes one or zero, z-score conversion results in infinite values. To avoid infinite values in the z-scores, floor and ceiling hit and false alarm rates were adjusted to *1–(1/2N)* when they were 1, and to *1/(2N)* when they were 0, where *N* is the number of trials on which the proportion is based (MacMillan & Creelman 2004). We corrected 11 values in the SH group, 1 value in the FL group and 1 value in the IFL group in this fashion.

[5] It should be noted that negative d' can sometimes be observed when participants exhibit behavior where they consistently translate the absence of signals to its opposite, and can arise due to task non-compliance, or random errors in performance.

and the parameters of signal detection theory (SDT) and the parameters of logistic regression are equal. Therefore, d' and c can be analyzed by using the log odds of hit rates and false alarm rates, where the logit is defined as the natural logarithm of the odds of a hit (or false alarm): $logit(p) = log(p/(1-p))$. When the logit transform is applied to hit and false alarm probabilities, the log-odds that the participant says "yes" to a signal and the log odds that they say "yes" to noise can be used as dependent measures in an ANOVA and meet normality assumptions. The d' can then be calculated from logits as the difference between the logit of hits and the logit of false alarms, and c can be calculated as -1 times the logit false alarms (DeCarlo 1998: 187). We supplement this analysis with non-parametric statistics, which is another way to deal with the absence of normality assumptions (but which lack the inference to the parent population of participants).

In addition, as suggested by a reviewer, we also conducted an error analysis of the FL group where the mean probability of false alarm rates for the two categories was compared using a paired sample t-test. This analysis aimed to look for the unintended SH bias that was found by Lai (2015) in the FL group. We also compared the first and second halves of the test block in a paired sample t-test to examine the detection ability across the blocks—a developing bias should be evident as a block effect. The aim of this analysis was to see whether the participants used an explicit strategy of learning during the test session where the frequency of grammatical words was higher than the ungrammatical words.

3 Results

3.1 Descriptive/Non-Parametric Statistics

When participants cannot discriminate between grammatical and ungrammatical word forms at all, $P(H) = P(FA)$ and $d' = 0$. Inability to discriminate means having the same rate of saying "yes" to grammatical words as to ungrammatical words. As long as $P(H) \geq P(FA)$, d' must be greater than or equal to 0 (Macmillan & Creelman 2004).

d' results for the SH condition showed that ungrammatical words were detected with a mean sensitivity of 1.283 ($SD = 1.20$). As for the FL conditions, ungrammatical words were detected with a mean sensitivity of 0.216 ($SD = 0.26$) in FL and 0.242 (SD = 0.22) in IFL. The mean bias rates for each condition were always negative, which was expected as a result of the oddball paradigm (see Figure 4 for a visual comparison). Besides, the median scores of the groups show that each group's median score is descriptively above zero, thus, each group has shown detection ability. Only one participant in the SH group had negative d' (–0.042), whereas the number of participants who had negative d' in the FL groups was five in the FL condition and three in IFL condition. Descriptive statistics are summarized in Table 3.

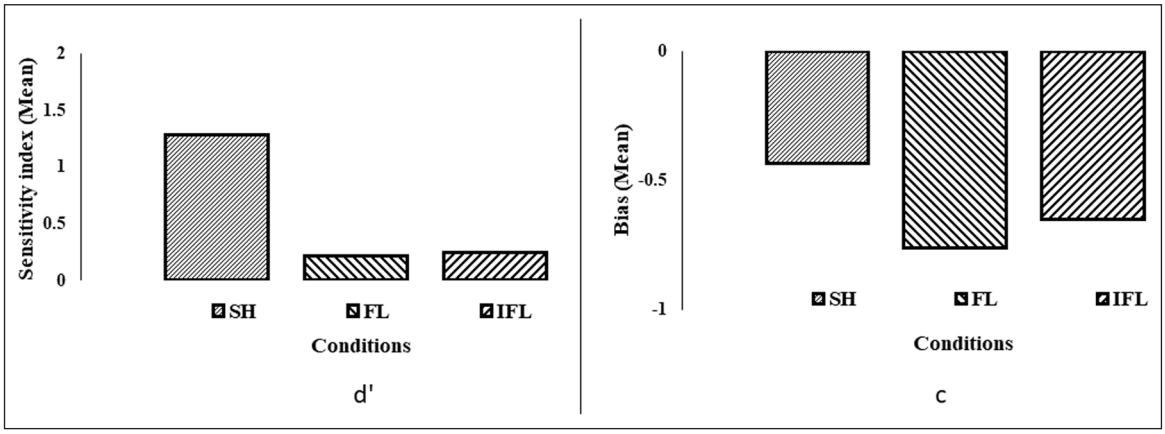

Figure 4: All Conditions: Group averages of sensitivity index rates of ungrammatical words (left panel), and bias rates of ungrammatical words (right panel).

Table 3: Descriptive statistics for the three experimental groups and normality test results along with Shapiro-Wilk values.

Descriptives	*SH*		*FL*		*IFL*	
	d'	*c*	*d'*	*c*	*d'*	*c*
Mean	1.283	−0.435	0.216	−0.763	0.242	−0.654
SE	0.246	0.081	0.054	0.077	0.045	0.080
Median	0.839	−0.528	0.271	−0.732	0.251	−0.657
Min	−0.042	−0.868	−0.386	−1.599	−0.275	−1.321
Max	4.272	0.683	0.607	−0.085	0.678	0.093
Shapiro-Wilk	W = 0.885 p = .010	W = 0.814 p < .001	W = 0.956 p = .364	W = 0.985 p = .963	W = 0.985 p = .970	W = 0.967 p = .584

To test whether the mean scores in each group were significantly different from zero, we conducted the non-parametric version of the one-sample t-test, the Wilcoxon Signed-Rank test. For the SH group, a Wilcoxon Signed-ranks test indicated that the sensitivity index was significantly different from zero, $W = 299$, $p < .001$. As for the FL groups, the sensitivity index was again significantly different from zero, $W = 258$, $p = .001$ for FL, and $W = 279$, $p < .001$ for the IFL group.

3.2 Inferential Statistics

After converting participants' hit rates and false alarm rates to their corresponding log-odds, we conducted the inferential statistical analysis. Specifically, we used the logits of hits and false alarms as dependent measures in a one-way ANOVA with three levels of the group, to test the hypothesis that the groups should differ.[6] The results of the one-way ANOVA (cf. Figure 5) showed that there was a significant difference between the groups for the logit transformed hit rate ($F(2,69) = 19.832$, $p < .001$, $\eta^2 = .365$, $1\text{-}\beta = .999$). Orthogonal contrasts were conducted for planned pairwise comparisons and revealed that the hit rate for the ungrammatical words was significantly lower for the FL and IFL groups, compared to the SH group ($t = 6.21$, $p < .001$). There was no statistically significant

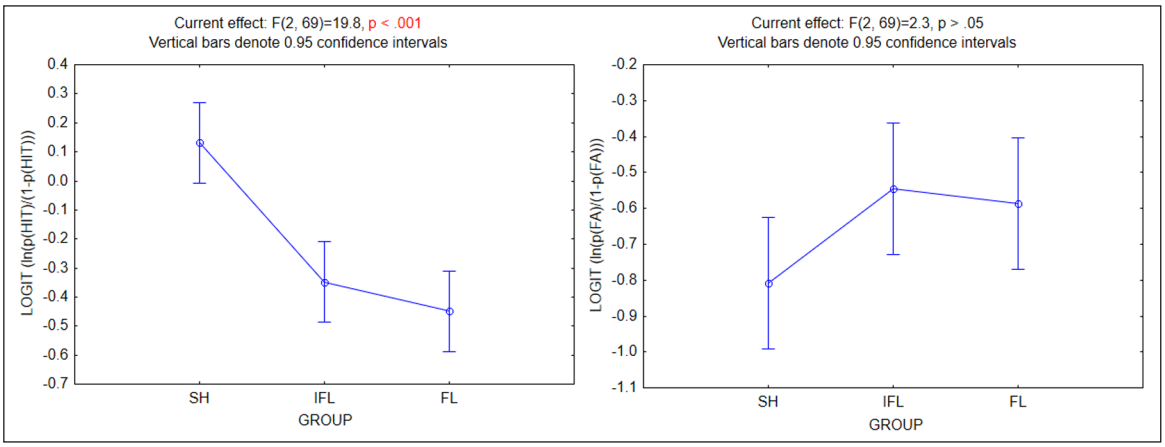

Figure 5: All Conditions: Group averages of logit transformed hit values (left panel), and false alarm values (right panel). Error bars indicate 95% CIs.

[6] All ANOVA effects are reported with the partial η^2 effect size measure and the t-tests with Cohen's *d*.

difference between the FL and IFL groups ($t = 1.02$, $p = .312$). This shows that participants who were trained with the sibilant harmony rule had a significantly higher hit rate than participants trained with FL. See Figure 5 left panel for a visual comparison of logit transformed hit values.

As for the false alarm rates, ANOVA results showed that the difference did not reach statistical significance, even though the trend was for the false alarm rate to be numerically lower in the SH group than in the FL groups ($F(2,69) = 2.370$, $p = .101$, $\eta^2 = .064$, $1\text{-}\beta = .463$). However, planned orthogonal contrasts revealed that the false alarm rate for the ungrammatical words was significantly higher for the FL and IFL ($t = 2.15$, $p = .035$) groups compared to the SH group. There was no significant difference in false alarms between the FL and IFL groups ($t = 0.32$, $p = .751$). The fact that the false alarm is numerically higher in the FL groups can be interpreted as follows: With minimal knowledge about the FL rule, the participants' task was to detect ungrammatical forms: thus, they were likely to get some more false alarms. On the other hand, the SH group had fewer false alarms because they were more confident about the rule. See Figure 5 right panel for a visual comparison of logit transformed false alarm values.

To aid interpretation of these logit results, we computed 95% confidence intervals around the mean logits and converted these values back into d', following DeCarlo (1998). The results showed that the mean d' for the SH condition was 2.164 with 95% CIs [3.031,1.297]. For the two FL condition, the mean d' was 0.319 with 95% CIs [0.492,0.145], and 0.454 with 95% CIs [0.636,0.272], respectively. This means that in each condition, the mean sensitivity level and their confidence intervals are above zero sensitivity, specifically, even though the FL and IFL groups had significantly lower sensitivity to ungrammatical forms than the SH group, the confidence intervals of the d' means (converted back from logit ANOVA) were higher than zero—in other words, the residual positive d' observed in the FL and IFL groups did not arise from chance guessing.

Our findings replicate Lai (2015)'s findings in that the attested SH pattern was significantly better learned by the participants, compared to the unattested FL patterns. However—and differently from Lai—we also observed a residual sensitivity to the FL rule in the FL and IFL groups, which contradicts Lai's previous conclusion that they should be unlearnable.[7]

3.3 Error analysis of the FL group

Before discussing our interpretation of the findings, we address the potential SH bias that was observed in the Lai (2015) study, by conducting an error analysis for the FL group. An SH bias in the current experimental context means that since a pattern that conforms to SH also conforms to FL, participants in the FL group, during their training of the FL patterns, might have developed a bias that makes them unwittingly learn the SH rule instead of the FL rule. One possible way to analyze this issue is to look at the errors participants made during the test, to see whether there is a pattern that supports a possible SH bias. In the context of the signal detection task, there are two errors: false alarms (signal was absent, but participant thought they detected it and reported so) and misses (signal was present, but the participant missed to report it). In terms of misses, since all violations of the FL pattern were at the same time violations of SH pattern, there is no way to differ-

[7] We also analyzed the variation within the stimuli (the position of the violating phoneme, k's position within the word and type of violation) across three conditions for the purpose of testing the null hypothesis that there should be no difference between the different types of stimuli. The results showed that the detection ability was not significantly affected by the position variations (all p values $>.05$).

entiate those errors. However, the examination of the false alarms would reveal whether there was an error pattern or not. In a hypothetical word, when the violation is in the middle of the word, the SH rule will be violated but not the FL rule. That is, during the testing, when a word in the form of [s.ʃ.s] or [ʃ.s.ʃ] was presented, an FL learner should not press the button to report a violation, but an SH learner should. If the FL learner presses the button, that raises the possibility that the FL learner induced SH instead of FL due to the words that conform to both rules in the training.

To this end, false alarms were coded as "FL-or-SH ([s.s.s] or [ʃ.ʃ.ʃ]), to reflect the words that conform to both rules, and as FL-only ([s.ʃ.s] or [ʃ.s.ʃ]) to reflect the words that follow only the FL pattern. The mean probability of false alarm rates for these two categories was compared using a paired sample t-test. The probability of false alarm for the FL-or-SH category was 0.22 (SD = 0.120), and 0.23 (SD = 0.103) for the FL-only category. Paired sample t-test results showed that there was no significant difference between the two categories, $t(23) = 0.53$, $p = 0.60$, $d = 0.109$. This demonstrates that the FL group did not have SH bias in that their error analysis showed that they did not have a significant preference for words that violate only the SH pattern. This reflects another difference in findings between our study and Lai (2015).

4 Discussion

The main objective of the current study was to replicate Lai's (2015) learning results in a different testing paradigm (oddball task). The results show that in each experimental condition, participants discriminated ungrammatical stimulus patterns with different levels of sensitivity. There are two main findings of this study: first of all, the sensitivity difference between the SH group and FL groups confirmed the previous findings that the difference in learnability is due to the computational complexity of the patterns. The second, and new, finding is that we have shown "residual" learning effects in participants trained on FL, an unnatural linguistic pattern.

In the SH condition, all the ungrammatical words were detected with a mean sensitivity higher than zero and biased at a negative mean rate; thus, *sensitivity* for ungrammatical words was better than zero sensitivity. This shows that participants were able to detect ungrammatical forms and acquired the rule based on the training data. The bias results showed that participants were conservative and biased to report no signal, which is also an expected consequence arising from the probability of the signal in an oddball design: fewer signals than no-signals are known to lead to negative bias (Eschman, St. James, Schneider & Zuccolotto 2005; Hilgard, Weinberg, Hajcak Proudfit, & Bartholow 2014).

An interesting anecdotal observation is that the rule learning in the SH participants was highly implicit. After the test session, participants were informally asked what they thought the rule was, but most participants who showed good detection ability nevertheless reported that they had no idea what the rule was, or they reported something wrong (e.g. that the rule was related to the vowels). This shows that the rule learning was implicit and not available to conscious reflection, as would be expected for an innately guided learning mechanism.

Positive d', as in the SH condition, was also observed in the other two experimental conditions (FL and IFL), although at much lower rates. From the formal language theory perspective, the learnability difference between the SH group and FL groups can be explained by the computational complexity of different subregular classes, namely the size of the window of segments over which the restriction is regulated in SLk or SPk languages. While the pattern present in the data can be learned with an SP $k = 2$ learner, an SL learner would require the window to be $k = 7$. Since the FL grammar must include position information, k will have to be larger for FL/IFL than SL and SP; thus, an

FL/IFL learner would require *k* to be at least 7. As pointed out by an anonymous reviewer, more data, time, and memory are necessary to accurately learn the pattern as *k* increases. Another point that needs to be noted learning is possible but less successful with a larger *k* due to performance factors like a limitation in short-term memory. Although the SH pattern can be learned with SP2 or SP4 grammars, since SP4 will need more memory, it is less memory-efficient for the learner because there are a lot more SP4 factors to consider than SP2 factors.

Participants in the FL groups were able, to some degree, to utilize the training part of the experiment to help them judge the grammaticality of the incoming stimuli. This finding was unexpected in the current study, so we do not have a specific account of the nature of this learning, apart from the fact that it is observed. In the following paragraphs, we speculate about the possible explanations of the residual learning effect observed in the FL groups. After showing evidence against an unintended SH bias and *on-demand* learning strategy, we will argue that residual learning in FL groups is due to general cognitive problem-solving abilities.

As discussed above, one possible explanation of this residual sensitivity in the FL groups might originate from the unintended SH bias: the idea that ambiguous stimuli that conform to both SH and FL rules helped FL learners to learn SH rule, as discussed in Lai (2015). However, our error analysis conducted for the FL group demonstrated that participants showed no differences between items that adhere to FL only vs. items that adhere to both SH and FL. In other words, our participants in the FL group did not show any evidence for the unintended SH bias. As for the learning observed in the IFL group, since the training stimuli in the IFL condition can be interpreted as a sibilant disharmony rule where each neighboring sibilant was disharmonic, the residual sensitivity levels in this group can be explained by referring to the learning of this pattern. This possibility was also discussed by Lai (2015). Nevertheless, we opt for the simpler explanation that the residual learning in the IFL and FL groups is due to non-linguistic general learning mechanisms.

Another possible explanation of the residual learning effect in FL groups relates the question of whether the participants in the unattested FL conditions might have been using an "on-demand" learning strategy (raised by an anonymous reviewer). Since the grammatical words were more frequent than ungrammatical words during the test session, participants could exploit the frequency statistics of the words and develop an idea about the pattern throughout the experiment. It is possible that the IFL/FL participants showed some signs of learning because they learned from the test items, which primarily followed IFL/FL patterns. To examine this, we compared the sensitivity index from the first and second half of the test phase in each group. If the IFL/FL participants used an online learning strategy, then their learning would steadily increase by the end of the second block. However, there was no difference between the first and second half of the test phase in terms of detection ability in FL groups as well as in the SH group (all p values $>.05$). These results demonstrate that FL learners did not use a strategy that would have led to better performance over time.

The fact that participants in the FL groups showed some sensitivity with d' values higher than zero seems to contradict the strong Subregular Hypothesis's learnability claims. However, we interpret it differently. First, note that the highly significant interaction between group and sensitivity to ungrammatical words shown in Figure 5 demonstrates a statistical difference between SH learning over FL learning. The Subregular theory makes discrete predictions about learnability, but the experimental data that support it are statistical. Second, we suggest that the residual learning effect is an artifact of the laboratory learning situation. As a reviewer pointed out, participants clearly can use general intelligence to solve language problems (e.g. crossword puzzles), and we cannot prevent

participants from trying to "solve the puzzle." Thus, the residual learning effect could simply be the result of general problem-solving strategies – similar to domain-general learning such as relying on the saliency of word edges (Endress & Mehler 2010). A similar conclusion was reached in the fMRI study by Musso et al. (2003), who trained participants to learn both linguistically attested rules in a language unknown to the participants, as well as linguistically unattested rules, violating principles of Universal Grammar. Although participants were able to behaviorally demonstrate in-laboratory learning of both rule types, only the UG-consistent syntactic rules activated Broca's area. We speculate that brain substrates relevant for linguistically attested rules like SH would similarly show different activation patterns compared to brain regions responsible for general problem solving and FL-rule learning.

Thus, it seems that the learning mechanisms for linguistic patterns are distinct from those of non-linguistic auditory or visual patterns. By default, human learners use domain-specific linguistic mechanisms (like subregular constraints) to learn artificial (but UG-grammatical) patterns in laboratory settings. When this constrained learning fails, they may rely on other learning mechanisms to solve the problem at hand, but those other mechanisms appear to not lead to fully successful learning in the linguistic domain. Nevertheless, we acknowledge that the assumption that domain-general mechanisms do not lead to successful learning compared to linguistic mechanisms must be examined in future research.

We suggest that the greater sensitivity to the SH pattern can be explained by the hypothesis of innate linguistic factors operative during learning, added on top of general psychological learning/problem-solving mechanisms. The Subregular Hypothesis can be thought of as an example of a domain-specific constraint on induction, such that patterns that are attested in human languages are channeled to language-specific learning modules.

5 Conclusion

In this paper, we compared the relative learnability of two long-distance harmony patterns (Sibilant Harmony vs. First-Last Assimilation) that differ typologically (attested vs. unattested) and computationally (Strictly Piecewise vs. Non-Counting). We proposed that abstract lab-induced rules are quickly translated into processing routines that generate real-time phonotactic predictions during auditory processing, and that this processing system is instrumental in pattern learning. This was supported by experimental results showing that adult learners prefer certain phonological patterns or distributions over others. These results substantiate the claims of the Subregular Hypothesis that a dedicated phonological module is active during real-time phonological parsing and to a significant extent constrains the learnability of specific phonotactic patterns. The fact that participants in the unattested FL groups showed a weak learning effect demonstrates that performance factors can mask the predictions of the Subregular Hypothesis.

Additional Files

The additional files for this article can be found as follows:

- **Appendix A.** Training Stimuli. DOI: https://doi.org/10.5334/gjgl.892.s1
- **Appendix B.** Test Stimuli. DOI: https://doi.org/10.5334/gjgl.892.s2

Acknowledgements

We would like to thank all colleagues who helped us improve this article. Special thanks go to Jeffrey Heinz and William Idsardi for their invaluable comments and suggestions, Regine Lai for providing the stimuli, and Ryan Rhodes for his thoughtful comments. Versions or parts of this paper were presented at different conferences and seminars; The 41st Annual

Penn Linguistics Conference (PLC41), 35th West Coast Conference on Formal Linguistics (WWCFL35), University of Delaware Experimental Group Meeting. We are grateful to audiences at those events, especially to the anonymous reviewers of PLC41 and WCCFL35. We also thank the anonymous reviewers for their constructive criticism. Finally, we thank the anonymous participants in our studies who volunteered their time and effort.

Competing Interests

The authors have no competing interests to declare.

References

Applegate, Richard. B. 1972. *Ineseno-Chumash grammar*. Berkeley, CA: University of California dissertation.

Archangeli, Diana & Douglas Pulleyblank. 2007. Harmony. In Paul de Lacy (Ed.), *The Cambridge Handbook of Phonology*, (Cambridge Handbooks in Language and Linguistics 353–378). Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511486371.016

Aslin, Richard N., Jenny R. Saffran & Elissa L. Newport. 1998. Computation of conditional probability statistics by 8-month-old infants. *Psychological Science* 9(4). 321–324. DOI: https://doi.org/10.1111/1467-9280.00063

Chambers, Kyle E., Kristine H. Onishi & Cynthia Fisher. 2003. Infants learn phonotactic regularities from brief auditory experience. *Cognition* 87(2). B69–B77. DOI: https://doi.org/10.1016/s0010-0277(02)00233-0

Chandlee, Jane. 2014. *Strictly local phonological processes*. Newark, DE: University of Delaware dissertation.

Chomsky, Noam. 1956. Three models for the description of language. *IRE Transactions on Information Theory* 2(3). 113–124. DOI: https://doi.org/10.1109/TIT.1956.1056813

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.

Cristia, Alejandrina. 2018. Can infants learn phonology in the lab? A meta-analytic answer. *Cognition* 170. 312–327. DOI: https://doi.org/10.1016/j.cognition.2017.09.016

DeCarlo, Lawrence T. 1998. Signal detection theory and generalized linear models. *Psychological Methods* 3(2). 186–205. DOI: https://doi.org/10.1037/1082-989X.3.2.186

Dell, Gary S., Kristopher D. Reed, David R. Adams & Antje S. Meyer. 2000. Speech errors, phonotactic constraints, and implicit learning: a study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26(6). 1355–1367. DOI: https://doi.org/10.1037/0278-7393.26.6.1355

Dettweiler, Stephen H. 2000. Vowel harmony and neutral vowels in C'Lela. *The Journal of West African Languages* 28(1). 3.

Edlefsen, Matt, Dylan Leeman, Nathan Myers, Nathaniel Smith, Molly Visscher & David Wellcome. 2008. Deciding strictly local (SL) languages. In *Proceedings of the midstates conference for undergraduate research in computer science and mathematics* 6. 6–75.

Endress, Ansgar D. & Jacques Mehler. 2010. Perceptual constraints in phonotactic learning. *Journal of Experimental Psychology. Human Perception and Performance* 36(1). 235–250. DOI: https://doi.org/10.1037/a0017164

Eschman, Amy, James St James, Walter Schneider & Anthony Zuccolotto. 2005. PsychMate: Providing psychology majors the tools to do real experiments and learn empirical methods. *Behavior Research Methods* 37(2). 301–311. DOI: https://doi.org/10.3758/BF03192698

Finley, Sara. 2009. Morphemic harmony as featural correspondence. *Lingua* 119(3). 478–501. DOI: https://doi.org/10.1016/j.lingua.2008.09.009

Finley, Sara. 2011. The privileged status of locality in consonant harmony. *Journal of Memory and Language* 65(1). 74–83. DOI: https://doi.org/10.1016/j.jml.2011.02.006

Finley, Sara. 2012a. Learning unattested languages. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* 34. 1536–1541. Austin, TX.

Finley, Sara. 2012b. Testing the limits of long-distance learning: Learning beyond a three-segment window. *Cognitive Science* 36(4). 740–756. DOI: https://doi.org/10.1111/j.1551-6709.2011.01227.x

Finley, Sara. 2015. Learning nonadjacent dependencies in phonology: Transparent vowels in vowel harmony. *Language* 91(1). 48–72. DOI: https://doi.org/10.1353/lan.2015.0010

Finley, Sara. 2017. Locality and harmony: Perspectives from artificial grammar learning. *Linguistics and Language Compass* 11(1). 1–16. DOI: https://doi.org/10.1111/lnc3.12233

Finley, Sara & William Badecker. 2009a. Artificial language learning and feature-based generalization. *Journal of Memory and Language* 61(3). 423–437. DOI: https://doi.org/10.1016/j.jml.2009.05.002

Finley, Sara & William Badecker. 2009b. Right-to-left biases for vowel harmony: Evidence from artificial grammar. In *Proceedings of the 38th North East linguistic society annual meeting* 1. 269–282.

Goldrick, Matthew. 2004. Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language* 51(4). 586–603. DOI: https://doi.org/10.1016/j.jml.2004.07.004

Green, David M. & John A. Swets. 1966. *Signal detection theory and psychophysics*. New York: Wiley.

Hale, Mark & Charles Reiss. 2000. "Substance abuse" and "dysfunctionalism": current trends in phonology. *Linguistic Inquiry* 31(1). 157–169. DOI: https://doi.org/10.1162/002438900554334

Hansen, Kenneth C. & Lesley E. Hansen. 1969. Pintupi phonology. *Oceanic Linguistics* 8(2). 153–170. DOI: https://doi.org/10.2307/3622818

Harrison, Michael A. 1978. *Introduction to formal language theory*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39(3). 379–440. DOI: https://doi.org/10.1162/ling.2008.39.3.379

Heinz, Jeffrey. 2010. Learning long-distance phonotactics. *Linguistic Inquiry* 41(4). 623–661. DOI: https://doi.org/10.1162/LING_a_00015

Heinz, Jeffrey. 2018. The computational nature of phonological generalizations. *Phonological Typology: Phonetics and Phonology*, 126–195. DOI: https://doi.org/10.1515/9783110451931-005

Heinz, Jeffrey, Chetan Rawal, and Herbert G. Tanner. 2011. Tier-based strictly local constraints for phonology. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume* 2. 58–64. Association for Computational Linguistics.

Heinz, Jeffrey & James Rogers. 2013. Learning subregular classes of languages with factored deterministic automata. In Andras Kornai & Marco Kuhlmann (eds.), *Proceedings of the 13th Meeting on the Mathematics of Language (MoL 13)*, 64–71, Sofia, Bulgaria: Association for Computational Linguistics.

Heinz, Jeffrey & William Idsardi. 2013. What complexity differences reveal about domains in language. *Topics in Cognitive Science* 5(1). 111–131. DOI: https://doi.org/10.1111/tops.12000

Hilgard, Joseph, Anna Weinberg, Greg Hajcak Proudfit & Bruce D. Bartholow. 2014. The negativity bias in affective picture processing depends on top-down and bottom-up

motivational significance. *Emotion* 14(5). 940–949. DOI: https://doi.org/10.1037/a0036791

Hopcroft, John E., Rajeev Motwani & Jeffrey D. Ullman. 2006. *Introduction to automata theory, Languages, and Computation*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.

Jardine, Adam. 2016. *Locality and non-linear representations in tonal phonology*. Newark, DE: University of Delaware dissertation.

Johnson, Douglas C. 1972. *Formal aspects of phonological description*. The Hague: Mouton. DOI: https://doi.org/10.1515/9783110876000

Kaplan, Ronald. M., & Martin Kay. 1994. Regular models of phonological rule systems. *Computational. Linguistics* 20(3). 331–378.

Keating, Pat. 2004. *D-prime (signal detection) analysis*. Retrieved from http://phonetics.linguistics.ucla.edu/facilities/statistics/dprime.htm.

Kisseberth, Charles. W. 1970. On the functional unity of phonological rules. *Linguistic Inquiry* 1(3). 291–306.

Kobele, Gregory M. 2006. *Generating copies: An investigation into structural identity in language and grammar*. Los Angeles, CA: University of California dissertation.

Koo, Hahn & Lydia Callahan. 2012. Tier-adjacency is not a necessary condition for learning phonotactic dependencies. *Language and Cognitive Processes* 27(10). 1425–1432. DOI: https://doi.org/10.1080/01690965.2011.603933

Lai, Regina. 2012. *Domain specificity in learning phonology*. Newark, DE: University of Delaware dissertation.

Lai, Regina. 2015. Learnable vs. unlearnable harmony patterns. *Linguistic Inquiry* 46(3). 425–451. DOI: https://doi.org/10.1162/LING_a_00188

Macmillan, Neil A. & C. Douglas Creelman. 2004. *Detection theory: A user's guide*. Psychology press.

Marcus, Gary F., Sugumaran Vijayan, S. Bandi Rao & Peter M. Vishton. 1999. Rule learning by seven-month-old infants. *Science* 283(5398). 77–80. DOI: https://doi.org/10.1126/science.283.5398.77

McNaughton, Robert & Seymour Papert. 1971. *Counter-free automata*. MIT Press.

Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25(1). 83–127. DOI: https://doi.org/10.1017/S0952675708001413

Musso, Mariacristina, Andrea Moro, Volkmar Glauche, Michel Rijntjes, Jürgen Reichenbach, Christian Büchel & Cornelius Weiller. 2003. Broca's area and the language instinct. *Nature Neuroscience* 6(7). 774–781. DOI: https://doi.org/10.1038/nn1077

Newport, Elissa L., & Richard N. Aslin. 2004. Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology* 48(2). 127–162. DOI: https://doi.org/10.1016/S0010-0285(03)00128-2

Ohala, John. J. 1993. The phonetics of sound change. *Historical Linguistics: Problems and Perspectives*, 237–278.

Onishi, Kristine H., Kyle E. Chambers & Cynthia Fisher. 2002. Learning phonotactic constraints from brief auditory experience. *Cognition* 83(1). B13–B23. DOI: https://doi.org/10.1016/S0010-0277(01)00165-2

Onnis, Luca, Padraic Monaghan, Korin Richmond & Nick Chater. 2005. Phonology impacts segmentation in online speech processing. *Journal of Memory and Language* 53(2). 225–237. DOI: https://doi.org/10.1016/j.jml.2005.02.011

Öttl, Birgit, Gerhard Jäger & Barbara Kaup. 2015. Does formal complexity reflect cognitive complexity? Investigating aspects of the Chomsky hierarchy in an artificial

language learning study. *PloS One* 10(4). e0123059. DOI: https://doi.org/10.1371/journal.pone.0123059

Peperkamp, Sharon, Rozenn Le Calvez, Jean-Pierre Nadal & Emmanuel Dupoux. 2006. The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition* 101(3). B31–B41. DOI: https://doi.org/10.1016/j.cognition.2005.10.006

Pulleyblank, Douglas. 2002. Harmony drivers: No disagreement allowed. In *Annual Meeting of the Berkeley Linguistics Society* 28. 249–267. DOI: https://doi.org/10.3765/bls.v28i1.3841

Pycha, Anne, Pawel Nowak, Eurie Shin & Ryan Shosted. 2003. Phonological rule-learning and its implications for a theory of vowel harmony. In *Proceedings of the 22nd west coast conference on formal linguistics* 22. 423–435.

Rogers, James & Geoffrey K. Pullum. 2011. Aural pattern recognition experiments and the subregular hierarchy. *Journal of Logic, Language and Information* 20. 329–342. DOI: https://doi.org/10.1007/s10849-011-9140-2

Rogers, James, Jeffrey Heinz, Gil Bailey, Matt Edlefsen, Molly Visscher, David Wellcome & Sean Wibel. 2010. On Languages Piecewise Testable in the Strict Sense. In Christian Ebert, Gerhard Jäger & Jens Michaelis (eds.), *The Mathematics of Language* 6149. 255–265. Springer. DOI: https://doi.org/10.1007/978-3-642-14322-9_19

Rogers, James, Jeffrey Heinz, Margaret Fero, Jeremy Hurst, Dakotah Lambert & Sean Wibel. 2013. Cognitive and sub-regular complexity. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8036 LNCS*, 90–108. DOI: https://doi.org/10.1007/978-3-642-39998-5_6

Sapir, Edward & Harry Hoijer. 1967. *The phonology and morphology of the Navaho language* 50. University of California Press.

Shieber, Stuart M. 1985. Using restriction to extend parsing algorithms for complex-feature-based formalisms. In *Proceedings of the 23rd annual meeting on Association for Computational Linguistics*, 145–152. DOI: https://doi.org/10.3115/981210.981228

Smolensky, Paul & Alan Prince. 1993. Optimality theory: Constraint interaction in generative grammar. *Optimality Theory in Phonology* 3.

Stanislaw, Harold & Natasha Todorov. 1999. Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers* 31(I). 137–149. DOI: https://doi.org/10.3758/BF03207704

Wilson, Colin. 2003. Experimental investigation of phonological naturalness. In *Proceedings of the 22nd west coast conference on formal linguistics*, 22. 533–546.

Zalcstein, Yechezkel. 1972. Locally testable languages. *Journal of Computer and System Sciences* 6(2). 151–167. DOI: https://doi.org/10.1016/S0022-0000(72)80020-5

How to cite this article: Avcu, Enes and Arild Hestvik 2020. Unlearnable phonotactics. *Glossa: a journal of general linguistics* 5(1): 56.1–22. DOI: https://doi.org/10.5334/gjgl.892

Submitted: 08 January 2019 **Accepted:** 20 March 2020 **Published:** 12 June 2020

]u[*Glossa: a journal of general linguistics* is a peer-reviewed open access journal published by Ubiquity Press. **OPEN ACCESS**