A challenge for tier-based strict locality from Uyghur backness harmony

Connor Mayer and Travis Major

UCLA, Los Angeles CA 90095, USA
connormayer@ucla.edu

Abstract. In this paper we describe the process of backness harmony in Uyghur, where suffix forms are determined first from the backness of certain vowels in the stem, or, if no such vowels are present, from the backness of dorsals in the stem. We show that this pattern cannot be captured by a tier-based strictly local (TSL) language. This is problematic for the weak subregular hypothesis, which claims that all segmental phonological stringsets are TSL languages. Next, we consider an alternative phonological analysis that is compatible with a TSL representation, but empirically unsupported. Finally, we consider the possibility that Uyghur backness harmony might be a lexicalized pattern, and find some suggestive evidence in support of this. This alternative appears to be the most likely way in which Uyghur backness harmony might, in principle, turn out to be compatible with the hypothesis that TSL languages provide an upper bound on phonological learnability.

Keywords: Uyghur, vowel harmony, backness harmony, subregular hierarchy, subregular hypothesis, formal complexity, phonology, TSL

1 Introduction

Researchers in computational linguistics propose that insights from theories of computation can guide how we study linguistic systems and what predictions we make about the structures of natural language (e.g. [21]). Hypothesizing that some aspect of language is bound by a particular computational structure has the potential to capture the wide variety of patterns seen across languages, while simultaneously constraining the types of patterns we should expect.

It is commonly accepted that phonological processes are regular: that is, they can be computed by regular grammars/automata (e.g. [28],[29]). This applies to both phonological stringsets (the properties of surface strings that may be characterized by phonotactic and markedness constraints) and phonological maps (the relationship between underlying and surface forms).

A stronger claim is that all phonological stringsets are tier-based strictly local (TSL) languages, which are subregular. That is, valid stringsets can be expressed as prohibitions on substrings, but these substrings may belong to "tiers" which contain only some subset of the segments in a language [23]. This is referred to as the *weak subregular hypothesis* [21].[1]

There are several existing counterexamples against TSL as an upper bound for phonological complexity. Some suprasegmental patterns have been identified as being outside of TSL, such as culminative quantity-sensitive stress rules [4] and circumambient patterns like unbounded tone plateauing [25]. A handful of segmental patterns that cannot be generated by single TSL grammars are described by McMullin [34]. McMullin claims that some of these exceptions can be captured using the intersection of multiple TSL grammars. However, the more complex patterns require a more powerful system, such as using an Optimality Theory account with constraints based on violations of TSL grammars. de Santo & Graf have formalized the intersection of multiple TSL languages as multi-tier strictly local (MTSL) languages, and proposed an extension of TSL, structure-sensitive TSL (SS-TSL), that allows the more problematic patterns described by McMullin to be captured [13]. Graf has also defined an extension of strictly piecewise (SP) grammars, interval-based SP (IBSP) grammars, which introduces domain restrictions to SP grammars and allows the problematic suprasegmental patterns described above to be captured [18].

This paper provides new data on a phonological process that is beyond the capacity of TSL grammars, providing a counterexample to the weak subregular hypothesis: backness harmony in Uyghur. This pattern is of interest for several reasons. First, it is a *segmental* process that cannot be generated by TSL grammars. These patterns are less common than suprasegmental patterns [25]. Furthermore, this pattern is significantly more complex than any segmental pattern previously discussed: Uyghur backness harmony cannot be captured by *any* of the classes that have been investigated in subregular phonology, with the exception of the overly powerful star-free languages. This makes it a particularly difficult case for anything but the weakest versions of the subregular hypothesis.

The paper is organized as follows. Section 2 will give a brief description of the properties of TSL languages. Section 3 will outline the characterization of backness harmony in Uyghur presented in the literature, and Section 4 will show that this pattern cannot be generated by grammars in any of the classes that have been previously investigated in subregular phonology. We briefly sketch an enhancement similar to SS-TSL that is able to capture this pattern, but leave its elaboration and implications for future research.

Given this data, it is either the case that previously considered subregular languages are insufficient to capture all phonological stringsets, or that another analysis for this phenomenon must be adopted. In Section 5 we describe an alternative characterization of Uyghur backness harmony that is compatible with a TSL analysis. We then present original experimental data suggesting that this analysis lacks empirical support. With this result in mind, Section 6 shows

[1] The *strong subregular hypothesis* claims that phonological stringsets are either strictly local (SL) or strictly piecewise (SP) languages [20]. Some autosegmental processes like stress have been claimed to be fully regular (e.g. [17]), though there is debate on whether alternative analyses are possible [20].

that Uyghur backness harmony exhibits many of the characteristics that are frequently taken as evidence of a lexicalized pattern. We elaborate on the implications of this in Section 7. If the description of Uyghur backness harmony in the literature is correct, then the weak subregular hypothesis is false. If this conclusion is to be avoided, then the most likely alternative to this description appears to be that the pattern is in fact lexicalized.

2 Tier-based strictly local languages

Tier-based strictly local languages fall within the subregular hierarchy, shown in Figure 1. Researchers have identified a variety of subregular language classes, and established their mathematical properties and the relationships between them (e.g. [44],[14],[45],[38]). We will not go into detail here about the subregular hierarchy in general, but many excellent descriptions and applications can be found elsewhere (e.g. [42],[41],[21],[48],[49]).

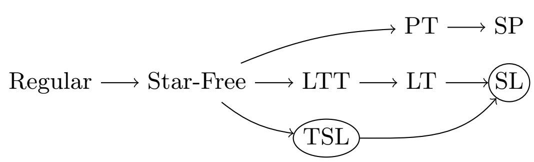

Fig. 1. The subregular hierarchy. Classes we discuss in this section are circled.

The class of TSL languages properly contains the class of strictly local (SL) languages and is properly contained within the class of star-free languages. It is incomparable with other subregular classes [23]. We first describe the properties of SL languages, and then examine how TSL languages expand on these.

Σ represents an alphabet. We will use the alphabet $\Sigma = \{a, b, c\}$ throughout the examples in this section. The symbols $\rtimes$ and $\ltimes$ are initial and final markers respectively, which are not in Σ. We will occasionally omit these for readability. The k-factors of a string $w \in \{\rtimes\} \cdot \Sigma^* \cdot \{\ltimes\}$ are defined as all substrings of w that are of length k, where $A \cdot B = \{ab \mid a \in A, b \in B\}$. A string u is a substring of a string w if $w = xuy$ for some strings $x, y \in \Sigma^*$. We can define a function $F_k(w)$ that returns the set of k-factors of w:

$$F_k(w) = \{u \mid u \text{ is a } k\text{-factor of } w\} \tag{1}$$

For example, $F_2(\rtimes ababac \ltimes) = \{\rtimes a, ab, ba, ac, c\ltimes\}$.

A strictly k-local grammar consists of a finite set of k-factors taken from $(\{\rtimes, \ltimes\} \cup \Sigma)^k$, which describe illicit substrings.[2] A string $w \in \{\rtimes\} \cdot \Sigma^* \cdot \{\ltimes\}$ is well formed with respect to a k-SL grammar G iff $F_k(w) \cap G = \emptyset$, i.e. if it

[2] These can equivalently be formulated as licit substrings.

contains no illicit substrings. A language L is SL iff there is some $k \in \mathbb{N}$ such that L can be generated by a strictly k-local grammar.

For example, suppose we want to define a language that prohibits all strings in which b is immediately followed by c. We could define a strictly 2-local grammar $G = \{bc\}$, which rules out strings such as $w_1 = ababca$, because $F_2(w_1) \cap G = \{bc\}$, but permits strings such as $w_2 = abacba$, because $F_2(w_2) \cap G = \emptyset$.

TSL grammars differ from SL grammars in that they are defined over a tier $T \subseteq \Sigma$ [23]. Only the segments on this tier are considered when checking for illicit k-factors. Formally, the representation of a string on a tier is generated by means of an erasing, or projection, function E_T, which removes symbols from the string that are not in T:

$$E_T(\sigma_1 \cdots \sigma_n) = u_1 \cdots u_n \tag{2}$$

where $u_i = \sigma_i$ iff $\sigma_i \in T$ and $u_i = \lambda$ (the empty string) otherwise. A k-TSL grammar consists of a set of k-factors taken from $(\{\rtimes, \ltimes\} \cup T)^k$. A string $w \in \{\rtimes\} \cdot \Sigma^* \cdot \{\ltimes\}$ is well formed with regard to a TSL grammar G iff $F_k(E_T(w)) \cap G = \emptyset$, i.e. if it contains no illicit substrings when projected on T. A language L is TSL iff it can be generated by a strictly k-local grammar over T for some $k \in \mathbb{N}$.

Suppose we want to define a language where words cannot contain both b and c. SL grammars are unable to capture this. We may define a strictly k-local grammar G that contains the k-factor $ba^{k-2}c$, where a^{k-2} represents the symbol a repeated $k-2$ times. This factor will rule out words like $aba^{k-2}ca$, but not words like $aba^{k-1}ca$, because the window of length k over which the k-factors operate is too small to see both the b and the c. Increasing k will not help, since it is always possible to increase the number of intervening a's. This is the result of a general property of SL languages [42]:

Theorem 1 (Suffix substitution closure). *A language L is SL iff there is some $k \in \mathbb{N}$ such that if there is a string x of length $k-1$ and strings u_1, t_1, u_2, and t_2, such that $u_1xt_1 \in L$ and $u_2xt_2 \in L$ then $u_1xt_2 \in L$.*

In contrast, it is trivial for a TSL grammar to capture this by letting $T = \{b, c\}$ and $G = \{bc, cb\}$. Under this formulation, the number of intervening a's is irrelevant, because they are excluded from T.

3 Uyghur backness harmony

Uyghur is a southeastern Turkic language with SOV word order. It has roughly 10 million speakers in the Xinjiang Uyghur Autonomous Region in the People's Republic of China and neighboring regions such as Kazakhstan and Kyrgyzstan. It has a rich system of vowel and consonant harmony along several dimensions. We focus here on backness harmony, which requires suffix forms to agree in backness with vowels and certain consonants within a stem.

The Uyghur vowels are shown in Table 1. The vowels behave as front or back as specified in the table, with the exception of /i/ and /e/, which are transparent

Table 1. The Uyghur vowel system. Harmonizing vowels are in bold.

	Front		Back	
	Unrounded	Round	Unrounded	Round
High	i	**y**		**u**
Mid	e	**ø**		**o**
Low	**æ**		**a**	

Table 2. The harmonizing Uyghur dorsal consonants

	Front	Back
Voiceless	k	q
Voiced	g	ʁ

to harmony processes [30][47].[3] We refer to /i/ and /e/ as *transparent* vowels, and the remainder of the vowels as *harmonizing* vowels. A subset of the dorsal consonants, shown in Table 2, also participate in backness harmony, with velars patterning with front vowels, and uvulars patterning with back vowels.[4]

Native Uyghur stems tend to be harmonious with respect to backness. This is not an absolute requirement for stems, however, and disharmonious stems are especially common in loanwords (e.g. /pæmidur/ 'tomato'). Such stems play a particularly interesting role in harmony processes.

The segments of a large class of Uyghur suffixes are underlyingly unspecified for backness. These suffixes take on the back feature of the stems they attach to. We will use the locative case marker /-DA/ as a representative example throughout the paper, but similar patterns occur with many other suffixes.

The examples in Table 3 provide a representative characterization of the pattern. Each example has a corresponding description of the particular type of harmony it illustrates. We refer back to these examples throughout the paper.

The voicing alternation of the initial segment in the suffix is not important, but note crucially that the vowel changes from front to back depending on the stem. The process for determining the backness value of the suffix is as follows:

1. Match the backness of the final harmonizing vowel in the stem. In (4) the stem is treated as a back stem because the final harmonizing vowel /o/ is back, and in (3) the stem is treated as a front stem even though it contains both front and back vowels because the final harmonizing vowel /æ/ is front.

2. If the stem has no harmonizing vowels, find the final dorsal consonant in the stem and match its backness. Note that in (5), the stem is treated as front even though it has only transparent vowels because the stem contains /g/, while in (6) the stem is treated as back because of its uvulars.

[3] Note that these vowels are the only ones in the system that have no counterparts differing only in backness. Because /e/ only occurs in loanwords and as the result of certain phonological processes, we focus primarily on /i/ throughout the paper.

[4] The velar sounds /x/ and /ŋ/ do not harmonize.

Table 3. Examples of Uyghur backness harmony. The alternating suffix is indicated in bold, and the harmony triggers are underlined.

Form	Gloss	Harmony type	
aʁinæ-**dæ** friend-LOC	"on the friend"	Closest front vowel	(3)
qoichi-**da** shepherd-LOC	"on the shepherd"	Closest back vowel	(4)
gezit-**tæ** newspaper-LOC	"on the newspaper"	Closest front dorsal	(5)
qirʁiz-**da** Kyrgyz-LOC	"on the Kyrgyz"	Closest back dorsal	(6)
rak-**ta** shrimp-LOC	"on the shrimp"	Closest back vowel across front dorsal	(7)
mæʃq-**tæ** exercise-LOC	"on the exercise"	Closest front vowel across back dorsal	(8)

Table 4. Examples of stems with arbitrary backness specification. The alternating suffix is indicated in bold.

Form	Gloss	Harmony type	
it-**ta** dog-LOC	"on the dog"	No harmonizers, arbitrarily back	(9)
biz-**dæ** we-LOC	"on us"	No harmonizers, arbitrarily front	(10)

Harmonizing vowels always take precedence over harmonizing dorsals, as (7) and (8) show. In these examples the harmonizing vowel determines the form of the suffix, even though a dorsal with the opposite backness intervenes. The process of only falling back on consonants to determine stem backness when insufficient information from vowels is available is the cause of the difficulties for TSL, as we will see in the next section.

Words with no harmonizing vowels or dorsals are arbitrarily specified for backness.[5] This is shown in Table 4. Such stems are theoretically problematic, but we will set them aside for now and return to them in Section 6.

[5] There is a statistical tendency for such stems to be treated as back.

4 The formal complexity of Uyghur backness harmony

In this section we focus on the pattern involving harmonizing vowels and dorsals described above. Because the actual segmental content is not of crucial importance, we use a more abstract notation to simplify the specification of the grammars. V_f and V_b refer to the sets of front and back harmonizing stem vowels.

$$V_f = \{y, \text{ø}, \text{æ}\} \tag{11}$$

$$V_b = \{u, o, a\} \tag{12}$$

C_f and C_b refer to the sets of front and back harmonizing dorsal stem consonants.

$$C_f = \{k, g\} \tag{13}$$

$$C_b = \{q, \text{ʁ}\} \tag{14}$$

We use the symbols S_f and S_b to refer to the sets of front and back suffix forms.

These symbols comprise an alphabet $\Sigma_h = \{V_f, V_b, C_f, C_b, S_f, S_b\}$. We define a homomorphic mapping function $h : \Sigma^* \mapsto \Sigma_h^*$ that converts strings from the full Uyghur alphabet to the notation described above (i.e. it maps stem symbols individually according to the definitions in (11) to (14), entire suffixes to S_f or S_b, and all other sounds to the empty string λ).

Uyghur backness harmony can be characterized succinctly with the following regular expression, which picks out licit strings. The class of regular languages is closed under homomorphism.

$$(\Sigma_h^* V_f \overline{V_b}^* S_f)|(\Sigma_h^* V_b \overline{V_f}^* S_b)|(\overline{(V_f|V_b)}^* C_f C_f^* S_f)|(\overline{(V_f|V_b)}^* C_b C_b^* S_b) \tag{15}$$

Thus it is clear that this pattern is at most regular.

4.1 Challenges for TSL

In this section we will show that Uyghur backness harmony as described in (15) cannot be captured using TSL languages, but first we must comment on our notation. Although the set of regular languages is closed under relabeling, the set of SL (and hence TSL) languages is not. For example, the language $(ab)^*$ is SL, but its image under the relabeling $\{a \mapsto c, b \mapsto c\}$, $(cc)^*$, is not SL. To avoid an increase in generative capacity, we apply this relabeling to the *grammar* rather than the language. In other words, the relabeling is applied to the k-factors defined in the grammar, and the resulting grammar filters out candidate strings in the image of that relabeling. This provably results in no increase in generative capacity so long as the mapping is many-to-one, as it is here [1].

To deal with the vowel component, we can define a grammar over the tier

$$T_v = V_f \cup V_b \cup S_f \cup S_b \tag{16}$$

containing the following illicit 2-factors:

$$V_f S_b \tag{17}$$

$$V_b S_f \tag{18}$$

These factors rule out forms like *[aʁinæ-da] (cf. (3)) and *[qoichi-dæ] (cf. (4)).

Harmony with dorsals can be captured by defining a grammar over the tier

$$T_c = C_f \cup C_b \cup S_f \cup S_b \tag{19}$$

containing the following illicit 2-factors:

$$C_f S_b \tag{20}$$

$$C_b S_f \tag{21}$$

These factors rule out forms like *[gezit-ta] (cf. (5)) and *[qirʁiz-dæ] (cf. (6)). Thus 2-TSL grammars can capture the vowel and consonant patterns in isolation.

The difficulty arises when combining these two patterns into a single TSL grammar. Because harmonizing dorsal and vowel information must be considered simultaneously, we must define a grammar over a tier that contains both the relevant dorsals and vowels:

$$T = T_v \cup T_c \tag{22}$$

The grammar over T must be able to look to the beginning of the string to check for the presence of harmonizing vowels. We extend T to contain $\rtimes$ and use $C = C_f \cup C_b$ for the sake of brevity. Suppose we define k-factors over T for some fixed k of the following form:

$$V_b C^{k-2} S_f \tag{23}$$

$$V_f C^{k-2} S_b \tag{24}$$

$$\rtimes C^{k-3} C_b S_f \tag{25}$$

$$\rtimes C^{k-3} C_f S_b \tag{26}$$

(23) and (24) try to capture harmony with vowels, while (25) and (26) try to capture cases with only harmonizing dorsals. These cannot work for all possible forms. Consider the following word on T:

$$w = V_b C_f^{k-1} S_f \tag{27}$$

This word, which has a mismatch between the final vowel and suffix form, will be erroneously included because the number of post-vowel dorsal consonants is too large for the k-factors to see both the vowel and suffix form at the same time. Checking for the absence of harmonizing vowels is bounded by k under suffix substitution closure, and therefore a TSL grammar over a tier containing both harmonizing dorsals and vowels cannot capture this pattern for arbitrary values of k, placing it outside of TSL.

Another possibility for capturing this pattern is to use the intersection of the TSL languages on T_v and T_c defined in (16)-(18) and (19)-(21) respectively. The class of TSL languages is not closed under intersection, and the resulting language falls in the class of multi-tier strictly local languages (MTSL), which properly contains the class of TSL languages [13]. Even this more powerful formalism cannot capture this pattern. The difficulty arises from the fact that violations on T_c should be ignored unless neither V_f nor V_b appear in T_v. Consider again examples (7) and (8), which we repeat in Table 5 along with their tier-based representations.

Table 5. Examples of Uyghur backness harmony over intervening, conflicting dorsals. The alternating suffix is indicated in bold, and the harmony triggers are underlined.

Form	Gloss	T_v	T_c	Harmony type	
r<u>a</u>k-**ta** shrimp-LOC	"on the shrimp"	$V_b S_b$	$C_f S_b$	Closest back vowel across front dorsal	(28)
m<u>æ</u>ʃq-**tæ** exercise-LOC	"on the exercise"	$V_f S_f$	$C_b S_f$	Closest front vowel across back dorsal	(29)

Violations on one tier cannot be overlooked given the contents of another, so this grammar rules both (28) and (29) as illicit because they are ill-formed on T_c. It would rule them ill-formed on T_v if suffixes of the opposite backness were used. Thus Uyghur backness harmony is also not MTSL.

4.2 Challenges for other subregular languages

The previous section showed that neither TSL nor MTSL grammars can capture the pattern in (15). We focused on these classes because they have received the most consideration as possible subregular upper bounds for phonotactic complexity. In this section we will sketch the arguments that the other subregular classes of languages that have been applied to phonology, including more expressive extensions of TSL, do not contain this pattern. We do not provide formal definitions of these languages here, but refer the reader to previous work.

Uyghur backness harmony is not SS-TSL or SS-MTSL Structure sensitive tier-based strictly local (SS-TSL) languages generalize the tier-projection process used in TSL [13]. TSL uses a 1-Input Strictly Local (1-ISL) projection, meaning that the projection function considers each segment in isolation (i.e. whether that segment is a member of T) [10]. SS-TSL generalizes this projection to a k-ISL projection, which means the projection function can consider a window of size k around the target segment. For example, we may define a SS-TSL grammar that will project a segment a to a tier only when it is immediately followed by segment b, but not otherwise. Structure sensitive multi-tier strictly local languages (SS-MTSL) are the intersection of multiple SS-TSL languages.

Intuitively, one might try to capture the Uyghur pattern by projecting harmonizing dorsals only when they are not preceded by a harmonizing vowel. It is easy to show using the suffix substitution closure property discussed at the end of Section 2 that cannot work for all forms. Assume a 2-SS-TSL grammar that includes the illicit 2-factor $C_b S_f$. Assume also that the projection function is k-ISL for some k, with the target segment falling into the final slot in the window (i.e. the context we consider is the target segment plus the preceding $k-1$ segments). The string $V_f C_b^k S_f$ will be excluded from this language even though it is a valid Uyghur word because the last of the k occurrences of C_b will be projected onto the tier. SS-MTSL fails for the same reason.

Uyghur backness harmony is not PT or SP Piecewise testable (PT) grammars are an extension of strictly piecewise (SP) grammars. SP grammars are similar to SL grammars but prohibit *subsequences* (i.e. precedence relations between segments) rather than substrings [45]. PT languages are the closure of SP languages under the Boolean operators $\wedge$ and $\neg$ [41]. Informally, these languages extend SP with the ability to *require* some subsequence be present in a string.

Even the basic vowel harmony pattern cannot be captured by a PT language. The intuition behind this is that the backness of suffixes is determined by the immediately preceding harmonizing vowel, but PT languages cannot precisely capture the order in which vowels occur. For example, $V_f V_b S_b$ and $V_b V_f S_b$ both contain the subsequences $V_f S_b$ and $V_b S_b$, but the first is a legal word while the second is not. We can show this more formally using the following theorem [41]:

Theorem 2 (k-Subsequence Invariance). *A language L is Piecewise Testable iff there is some $k \in \mathbb{N}$ such that for all strings x and y, if x and y contain the same set of subsequences of length k or less, then either $x \in L$ and $y \in L$ or $x \notin L$ and $y \notin L$.*

Consider the following pair of words for some $k \in \mathbb{N}$:

$$w_1 = (V_f V_b)^k S_b \tag{30}$$

$$w_2 = (V_b V_f)^k S_b \tag{31}$$

These words contain the same subsequences of length k or less,[6] but w_1 is a valid word while w_2 is not. Thus even the simplest subcase of Uyghur backness harmony is not PT, and since PT properly contains SP, it is also not SP.

Uyghur backness harmony is not LTT or LT Locally threshold testable (LTT) grammars are an extension of locally testable (LT) grammars. LT languages are the closure of the SL languages under the Boolean operators $\wedge$ and $\neg$ [42][41]. Informally, these languages extend SL with the ability to *require* some element be present in a string. LTT languages are the closure of LT languages under the first order logic operators $\forall$ and $\exists$, which quantify over position indices

[6] This can be shown by induction: both words contain the same subsequences when $k = 1$, and the subsequences added with each increase in k will be the k-subsequences generated by prepending V_f or V_b to all subsequences of length $k-1$.

[42][41]. Indices can be compared for equality and successorship. Informally, this extension allows LTT languages to count the number of occurrences of each k-factor up to a certain threshold.

It is simple to show that Uyghur backness harmony as described in (15) is outside of LTT by appealing to the following theorem [41]:

Theorem 3 (Local Threshold Test Invariance). *A language L is Locally Threshold Testable iff there is some $k \in \mathbb{N}$ and some threshold $t \in \mathbb{N}$ such that, for all strings x and y, if for any k-factor w, x and y contain the same number of occurrences of w or have at least t occurrences, then either $x \in L$ and $y \in L$ or $x \notin L$ and $y \notin L$.*

Consider the following two words for some $k \in \mathbb{N}$:

$$w_1 = (C_f)^{k-1} V_b (C_f)^{k-1} V_f (C_f)^{k-1} S_f \quad (32)$$

$$w_2 = (C_f)^{k-1} V_f (C_f)^{k-1} V_b (C_f)^{k-1} S_f \quad (33)$$

Both have the same number of occurrences of every k-factor, but w_1 is a valid Uyghur word while w_2 is not. Therefore Uyghur backness harmony is not LTT, and since LTT properly contains LT, it is also not LT.

Uyghur backness harmony is not IBSP Interval-based strictly piecewise (IBSP) grammars are an extension of SP grammars that allow k-subsequences to be defined over a particular interval, such as a word or a prosodic phrase [18]. The set of IBSP languages properly contains both TSL and SP languages, and is properly contained by the star-free languages. Uyghur backness harmony is a word-level process, and an IBSP grammar that is defined over words will encounter the same issues as the PT and SP grammars described above. We can think of no interval below the word that is able to avoid these problems, and so we conjecture that Uyghur backness harmony is not IBSP.

4.3 A formal lower bound for Uyghur backness harmony

The pattern in (15) can be captured by the non-counting (NC) or star-free languages, which are the most expressive subregular languages [35]. NC languages allow the use of the first order logic operators $\exists$ and $\forall$, which quantify over position indices in the string. Indices can be compared for equality, using the $\approx$ operator, and precedence, using the $<$ operator. Predicates over indices $P(x)$ are true if the symbol at index x is P. All of the language classes described above are properly contained by the class of NC languages.

The following expressions define a NC grammar that captures licit forms under Uyghur backness harmony.

$$\forall x[S_b(x) \Rightarrow \forall y[V_f(y) \Rightarrow \exists z[V_b(z) \wedge y < z < x]]] \quad (34)$$

$$\forall x[S_f(x) \Rightarrow \forall y[V_b(y) \Rightarrow \exists z[V_f(z) \wedge y < z < x]]] \quad (35)$$

$$\forall x[S_b(x) \wedge \neg\exists y[V_f(y) \vee V_b(y)] \Rightarrow \forall z[C_f(z) \Rightarrow \exists w[C_b(w) \wedge z < w < x]]] \quad (36)$$

$$\forall x[S_f(x) \wedge \neg\exists y[V_f(y) \vee V_b(y)] \Rightarrow \forall z[C_b(z) \Rightarrow \exists w[C_f(w) \wedge z < w < x]]] \quad (37)$$

The first two expressions require suffixes to match the backness of the final harmonizing stem vowel. The latter two require suffixes to match the backness of the final harmonizing stem dorsal if there are no harmonizing stem vowels.

Although further weakening the weak subregular hypothesis to include NC languages captures the data presented above, this is not a desirable result from the perspective of learnability. It has been shown that TSL languages are efficiently learnable in polynomial time from polynomial data [22][26][27], while NC languages are not [16]. This makes theories of phonological learning somewhat more problematic. Below we briefly sketch a proposal for a new subregular class that is less powerful than the NC class but still sufficient to capture this pattern.

4.4 OSS-TSL

A generalization of TSL, which we tentatively call output structure-sensitive tier-based strictly local (OSS-TSL), can capture Uyghur backness harmony.[7] SS-TSL generalizes the projection function of TSL from a 1-ISL map to a k-ISL map. The class of 1-ISL maps is identical to the class of 1-output strictly local (1-OSL) maps, meaning the TSL projection function could be equally characterized as a 1-OSL function [10]. We could thus generalize the projection mechanism to be k-OSL. This would allow the tier projection function to consider material already on the tier when choosing whether to project a segment from the input.

The pattern in (15) requires a 2-OSL projection function that behaves as follows: V_f, V_b, S_f, and S_b are always projected, while C_f and C_b are only projected if the previous symbol is not V_f or V_b. In other words, we stop adding dorsals to the tier as soon as we encounter a harmonizing vowel. 2-factors defined over this tier would simply check for backness mismatches between the suffix and the preceding symbol.

It is beyond the scope of this paper to consider the formal properties of OSS-TSL grammars and how widely applicable they will be in describing natural language phonology. We intend to investigate this in future research.

5 An analysis without transparent vowels

Given the uncommonness of segmental patterns that are as problematic for the weak subregular hypothesis as Uyghur backness harmony, it is worth investigating whether the characterization of the pattern presented above is correct. The issues this pattern poses for TSL representations hinge on backness being determined first from vowels, and then from consonants if the vowels prove insufficient. A possible alternative analysis that is compatible with a TSL representation is that Uyghur in fact has no transparent vowels. Rather, there are two different surface versions of /i/ and /e/ which are not reflected in the orthography or in past descriptions of the phonology, one of which is front and one of which is

[7] We also suggest that SS-TSL might be relabeled as input structure-sensitive tier-based strictly local (ISS-TSL).

back. We represent the back variants as /ɨ/ and /ɘ/ respectively, and refer to the pairs as /I/ and /E/ when backness is not important. Such an analysis makes the 2-factors defined over T_v in the previous section sufficient to capture Uyghur backness harmony: V_f now includes /i/ and /e/, and V_b includes /ɨ/ and /ɘ/, so no reference to consonants is necessary. This account is supported by historical evidence: Uyghur once had a distinction between the front and back unrounded vowels /i/-/ɯ/ and /e/-/ɤ/ as in Turkish, but these vowels have since collapsed into /i/ and /e/ [30].

Under this formulation, /I/ and /E/ are underlying specified as [+/- back]. This allows us to tidily capture forms like (9) and (10), which no longer need to be arbitrarily specified as front or back, but now select their suffix based on the quality of the final vowel, as below:

ɨt-**ta** "on the dog"
dog-LOC (38)

biz-**dæ** "on us"
we-LOC (39)

The generalization that suffixes tend to match the backness of the final harmonizing vowel in the stem, or, if these are lacking, the final dorsal, can be captured by cooccurrence restrictions: /I/ and /E/ must agree in backness with the nearest harmonizing vowel or dorsal, which gives the appearance of suffixes harmonizing with consonants. Thus we can reanalyze (4), (5), and (6) as below.

qoɵchɨ-**da** "on the shepherd"
shepherd-LOC (40)

gezit-**tæ** "on the newspaper"
newspaper-LOC (41)

qɨrʁɨz-**da** "on the Kyrgyz"
Kyrgyz-LOC (42)

In sum, this approach allows us to determine suffix backness by looking only at the final vowel in the stem, which is always specified for backness. This removes the need for a dorsal consonant tier, and allows this pattern to be captured easily by the TSL grammar over T_v described in the previous section.

There are two issues with this approach. The first is that there are a small number of stems in the language that still appear to follow a pattern where vowels are considered before consonants, as exemplified below:

tæstiq-**tæ** "on the sanction"
sanction-LOC (43)

Here we might expect the final /i/ of the stem to take on the backness of the adjacent uvular, but instead we see that this word takes a front suffix. Words like this are rare, but it is unclear how to achieve the proper backness of the final vowel. One solution is to stipulate that vowels are considered before consonants, which leads back to the problem we hoped to avoid. Another solution is to stipulate that the surface forms of /I/ and /E/ are identical to their underlying forms, and the tendency for these forms to agree in backness with nearby harmonizing vowels and dorsals is simply a coincidence.

The second issue is that there have been no studies looking at the phonetic realization of /i/ and /e/ in stems that differ in the backness of their suffixes. Positing two separate phonemes without such data for the sake of a more theoretically amenable analysis is rather ad hoc, although it has been done. Lindblad, for example, proposes such an analysis, where an underlying contrast between /i/ and /ɨ/ is neutralized by a post-lexical fronting process [30]. This is essentially unfalsifiable. In the following section we present original data from a small study investigating whether there is phonetic support for such an analysis.

5.1 An acoustic study of the transparent vowel /i/ in Uyghur

In the present study we restrict ourselves to stems containing neither harmonizing vowels nor dorsals. This is because coarticulatory processes, where the articulation of a sound is influenced by nearby sounds, are a confounding factor in trying to show that there is a phonological distinction between /i/ and /ɨ/ in Uyghur. Coarticulatory processes are common across languages, and though they often become phonologized, the presence of coarticulation does not necessarily provide evidence for a phonological process (e.g. [24],[36]). In other words, we cannot tell using measurements whether the vowels in (42), for example, are phonologically [ɨ] or simply a little backer because of the nearby uvulars.

Vowel-to-vowel coarticulation has been studied extensively (e.g. [3],[11],[12]). A finding that is relevant for the pattern described here is that languages vary in their patterns of V-to-V coarticulation: coarticulation tends to be greater when there is less risk of confusion between meaningfully distinct phonemes [32][31]. This suggests that, under the assumption that phonological /ɨ/ and /ɘ/ do not exist, /i/ and /e/ should be the most susceptible of the Uyghur vowels to coarticulation with nearby back sounds, since these are the only two that have no corresponding vowel differing only in backness.

Less has been said about the effects of uvular consonants on vowels, but studies of languages such as Cochabamba Quechua [15] and Ditidaht [46] show that uvulars produce a backing effect on nearby vowels (particularly front vowels).

These findings suggest that the vowels /i/ and /e/ should exhibit backing around back vowels and uvulars by phonetic coarticulatory processes, and hence acoustic evidence of such cannot be taken as proof of a phonological distinction. A domain that is free from this confound is the set of stems that have neither harmonizing vowels nor dorsal consonants. A significant difference in backness between the vowels in stems that take front suffixes and those that take back suffixes could not be attributed to phonetic coarticulation, and would be compelling evidence for a meaningful phonological contrast. Phonetic studies have been performed for a similar process in Hungarian [5][6], and the methodology here is similar to that employed by Blaho & Szeredi.

Methodology Two native speakers of Uyghur from the Urumchi region, one male and one female, read the words in Tables 6 and 7. All words are monosyllabic with only the transparent vowel /i/ and no dorsal consonants. Both speakers were presented with the same list of words. Words were removed if they were unfamiliar to the speaker or if they were produced with no vowel.[8] The speakers did not agree on the backness of all words when suffixes were added: bolded words in the tables highlight disagreements.

Tables 6 & 7. Word lists for speakers 1 and 2. Bolded forms indicate disagreements in stem backness between the speakers.

Front		**Back**	
/bil/	‘know’	/ʧiʃ/	‘tooth’
/bir/	‘one’	/dil/	‘heart’
/biz/	‘we’	**/mis/**	**‘copper’**
/din/	**‘religion’**	/pil/	‘elephant’
/iʃ/	‘work’	/sirt/	‘outside’
/ʤin/	**‘Djinn’**	/siz/	‘draw’
/min/	‘ride’	/til/	‘tongue’
/sir/	**‘brush’**	/tiz/	‘knee’
/siz/	‘you’		

Front		**Back**	
/bil/	‘know’	/ʧiʃ/	‘tooth’
/bir/	‘one’	/dil/	‘heart’
/biz/	‘we’	**/din/**	**‘religion’**
/min/	‘ride’	/it/	‘dog’
/mis/	**‘copper’**	**/ʤin/**	**‘Djinn’**
/siz/	‘you’	/lim/	‘beam’
		/pil/	‘elephant’
		/pir/	‘master’
		/sir/	**‘brush’**
		/sirt/	‘outside’
		/siz/	‘draw’
		/til/	‘tongue’
		/tiz/	‘knee’

The speakers produced the words in the carrier sentence

tursun hazir	________	*dɛdi*	
Tursun again	________	say.PAST	(44)
Tursun said	________	again.	

Words were elicited in two forms: with no harmonizing suffix (bare for nouns, and with the third person past tense suffix -*di* for verbs) and with a harmonizing suffix (the locative -*DA* for nouns, and the infinitive -*mAQ* for verbs, where Q alternates between /k/ and /q/ depending on backness). The purpose of eliciting words with a suffix was to confirm the predicted coarticulatory effect of a nearby back or front vowel.

The stem vowels were segmented by hand using Praat [7], and F1 and F2 were extracted at vowel midpoints using a script. We ran two linear mixed effects models in R [40] using the *nlme* package [39], with F1 and F2 as the dependent variables respectively. Stem backness (i.e. whether the stem takes a front or back

[8] Uyghur has a process of vowel lenition that can occur adjacent to voiceless consonants: e.g. speaker 1 produced /ʧiʃ/ and /iʃ/ with vowels while speaker 2 did not, and speaker 2 produced /it/ with a vowel while speaker 1 did not.

suffix) and the presence of a harmonizing suffix were the independent variables. Word and subject were random effects.

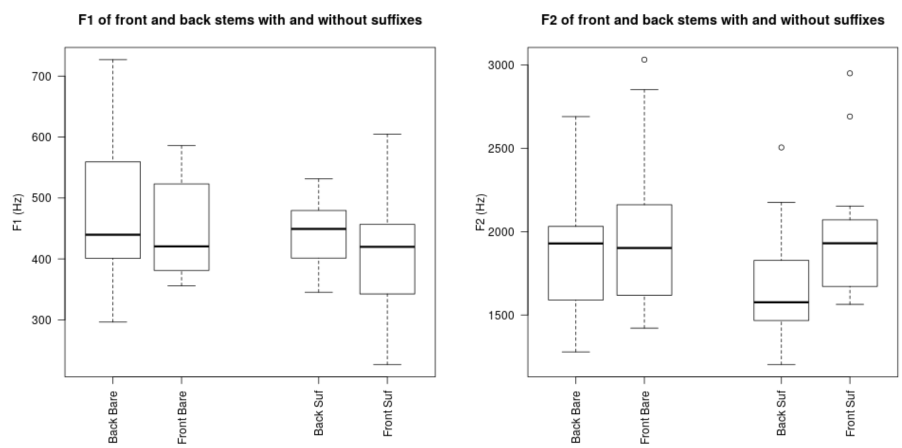

Fig. 2. Plots of F1 (left) and F2 (right)

Results The results are shown in Figure 2. There were no significant effects of stem backness or suffix presence on F1. Similarly, there was no significant effect of stem backness on F2, but there was a significant interaction between stem backness and the presence of a suffix on F2 ($\beta = 183.47$; $t = 3.65$; $p < 0.01$). This indicates that back vowels in the suffixes pulled /i/ back, lowering F2. These results provide no evidence that the vowel /i/ provides a cue for backness in stems with no harmonizing vowels or dorsals, though they do demonstrate a coarticulatory effect between /i/ and nearby back vowels.

Discussion The fact that nearby back vowels induce backing on /i/ suggests that /i/ is indeed susceptible to coarticulation with neighboring segments. Hence it is plausible that in stems containing harmonizing vowels or dorsals, information about these segments may be conveyed through the transparent vowels. The challenge for adopting this analysis is making a convincing case that this is a phonological pattern, not merely a phonetic one. The general results here do not support this analysis: stems containing only the vowel /i/ and non-dorsal consonants show no acoustic distinction based on the backness of their suffixes, and must still be arbitrarily specified as front or back. Although this analysis is able to circumvent the issues for a TSL representation, it does so with no gain in explanatory power, and at the cost of additional stipulation.

6 Uyghur backness harmony as a lexicalized process

The characterization of Uyghur backness harmony in Sections 3 and 4 is incompatible with the theory that all phonological stringsets are TSL languages. There is, however, evidence that backness harmony in Uyghur may be learned as a lexical or morphological process, rather than a phonological one, which is consistent with the idea that TSL provides an upper bound on phonological complexity.

Under a morphological account of backness harmony, all stems are specified lexically as taking either front or back suffixes (or, put in a slightly different way, speakers simply memorize which stems take which suffixes). This pattern is easily captured using a TSL grammar, and is consistent with the hypothesis that morphotactic processes are also maximally TSL [1].

Under this analysis, stems lacking harmonizing vowels or dorsals are treated identically to all other stems. What this approach sacrifices is the strong generalization that stems with harmonizing vowels and dorsals tend to take suffixes that agree in backness. Although such approaches run counter to the tendency in generative phonology to limit lexical specification, there is some motivation for adopting them in certain cases. We will present a brief summary of a particularly notorious case, and show that Uyghur satisfies the same motivating criteria.

Positing a morphological process over a phonological one is often based on the complexity of the phonological analysis required to capture the pattern, particularly regarding learnability. Such analyses often require underlying forms that differ substantially from any surface form, baroque interactions between independent processes (e.g. rules or constraints), and some way to capture inconsistent generalizations and variation within or between speakers. Examples include French liaison (e.g. [9]), Polish /o/-/u/ alternation [43], irregular English past tense morphology [2], and possessive prefixes in Odawa [8]. Such cases have two common themes. First, these processes typically originated as predictable and productive phonological patterns that were subsequently obscured by diachronic change. This led to a reanalysis by language learners, since insufficient evidence was available to reconstruct the original pattern. Second, in the absence of reliable structural cues, speakers tend to rely on statistical generalizations to determine the appropriate surface realization in unfamiliar cases.

We focus on Maori passives as a representative example. This was first raised as a challenge for phonological analysis by Hale [19] and has been written on extensively since (see [37] for an excellent overview). Table 8 (from [37]) shows a sample of Maori passive forms.

This pattern developed as the result of all word-final consonants being lost in unsuffixed forms, but maintained in medial position when the passive suffix /-ia/ is present. A phonological analysis must either make reference to properties of the stems that systematically predict particular suffix forms (which are not obvious) or assume the presence of word-final consonants underlyingly and a process of surface deletion. Even in the latter case, bizarre assumptions are often required to support this analysis, such as Hale's suggestion that there is an underlying /p/ at the end of certain forms that is never realized in *any* surface form [19]. There is also evidence that these passive forms have been analyzed as separate, competing

Table 8. Maori passives

Active	Passive	Gloss
/ɸera/	/ɸerahia/	‘to spread’
/oma/	/omakia/	‘to run’
/inu/	/inumia/	‘to drink’
/eke/	/ekeŋia/	‘to climb’
/tupu/	/tupuria/	‘to grow’
/aɸi/	/aɸitia/	‘to embrace’
/huna/	/hunaia/	‘to conceal’
/kata/	/kataina/	‘to laugh’
/ako/	/akona/	‘to teach’
/heke/	/hekea/	‘to descend’

morphemes by speakers, with /-tia/ coming to be preferred as the default but substantial free variation possible within and between speakers. Sophisticated statistical analyses also show that the suffix form can be predicted reasonably well from subtle properties of the stem, suggesting that speakers may be sensitive to statistical generalizations when choosing the appropriate suffix [37].

Uyghur backness harmony shares many of the properties of the Maori passive system. Disharmonic stems tend to occur mostly in loanwords and compounds [47], suggesting an increase in such stems as more foreign words entered the language. Similarly, as described earlier, historical Uyghur once had a distinction between the front and back vowels /i/-/ɯ/ and /e/-/ɤ/ that collapsed into /i/ and /e/, eliminating the backness contrast that would have determined the suffix of many of the problematic forms discussed here [30]. As demonstrated by the responses of the participants in the study in the previous section, there is also inter-speaker variation on which suffixes certain forms take. We are conducting corpus and experimental studies to evaluate the extent to which the process of backness harmony has been productively acquired by Uyghur speakers [33].

More evidence is needed to establish that Uyghur backness harmony is treated by speakers as a lexicalized process, but there are several points in favor of this account: it provides a consistent analysis, with the generalizations around harmonizing vowels and dorsal consonants reflecting statistical remnants of a past, more reliable stage of the language; it exhibits many of the properties of languages where similar processes have been claimed; and, it is consistent with the theory that TSL provides an approximately correct upper bound on phonological learnability [34]. Assuming that only phonological patterns that can be effectively learned will be effectively propagated, it is perhaps no coincidence that non-TSL patterns are so uncommon in the world's languages.

7 Discussion and conclusion

A motivation that is commonly put forward for studying phonological processes through the lens of formal complexity is that it allows a detailed characterization of the data that is agnostic to any particular theory. This in turn provides clear requirements for the expressive and restrictive capabilities of any theory leveled at the data (e.g. [42],[21]), and also suggests upper bounds on the complexity of patterns that can be effectively learned (e.g. [34]). In this paper we presented a pattern that is a challenge for current hypotheses about how complex phonological stringsets may be.

We first showed that the pattern of stem-suffix backness harmony in Uyghur cannot be generated by a tier-based strictly local grammar, nor by *any* of the subregular language classes previously applied to phonology. This is problematic for the weak subregular hypothesis, which claims that all phonological stringsets are maximally TSL. We then explored an alternative analysis that suggests the pattern is phonological and contained solely on the vowel tier, but rejected it for lack of empirical support. Finally, we showed that even though Uyghur surface strings are beyond TSL, there is some evidence that the pattern of backness harmony may be lexicalized. This is consistent with the idea of TSL as an upper bound on effective phonological learnability. More empirical data is needed to claim with certainty that this is the case, but we feel that this is a promising hypothesis.

We do not go into considerations of the merits of one phonological theory over another, but it is worth noting that the characterization of Uyghur backness harmony bears on this as well. The interaction of backness harmony with other phonological process in Uyghur has been put forward as evidence for a derivational theory of phonology [47], and the considerations discussed here bear on the validity of these claims.

Assuming that backness harmony in Uyghur is governed by the phonotactic knowledge of its speakers, the analysis presented in Sections 3 and 4 is incompatible with the weak subregular hypothesis. If, on the other hand, backness harmony is indeed a lexicalized process, this would be consistent with the idea that TSL languages provide an approximately correct upper bound on phonological learnability. In addition to the simple presentation of this data as a challenging case for subregular phonology, we hope that we have illustrated how theories of formal complexity can serve as useful conceptual tools in addition to those traditionally employed by phonologists.

Acknowledgements We would like to thank Tim Hunter, Kie Zuraw, Thomas Graf, two anonymous reviewers, and the attendees of the UCLA phonology seminar for their invaluable feedback. We would also like to thank our Uyghur consultants for sharing their language and culture. Without their generosity and time, none of this would have been possible.

References

1. Alëna Aksënova, Thomas Graf, and Sedigheh Moradi. Morphotactics as tier-based strictly local dependencies. *Proceedings of the 14th Annual SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 121–130, 2016.

2. Adam Albright and Bruce Hayes. Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition*, 90:119–161, 2003.
3. Peter J. Alfonso and Thomas Baer. Dynamics of vowel articulation. *Language and Speech*, 25:151–173, 1982.
4. Hyunah Baek. Computational representation of unbounded stress: tiers with structural features. *Proceedings of the 53rd Annual Meeting of the Chicago Linguistic Society*, page To appear, 2017.
5. Stefan Benus and Adamantios Gafos. Articulatory characteristics of hungarian 'transparent' vowels. *Journal of Phonetics*, 35:271–300, 2007.
6. Sylvia Blaho and Dániel Szeredi. Hungarian neutral vowels: A microcomparison. *Nordlyd*, 40:20–40, 2013.
7. Paul Boersma and David Weenink. Praat: doing phonetics by computer [Computer Program]. Version 6.0.37. http://www.praat.org/. Accessed: 2018-02-17.
8. Dustin Bowers. *A system for morphophonological learning and its implications for language change*. PhD thesis, UCLA, 2015.
9. Joan Bybee. Frequency effects on French liaison. In J. Bybee and P. Hopper, editors, *Frequency and the Emergence of Linguistic Structure*, pages 337–359. John Benjamins, Amsterdam, 2001.
10. Jane Chandlee. *Strictly local phonological processes*. PhD thesis, The University of Delaware, 2014.
11. John-Dongwok Choi and Patricia A. Keating. Vowel-to-vowel coarticulation in Slavic languages. *Journal of the Acoustical Society of America*, 88, 1990.
12. Jennifer Cole, Gary Linebaugh, Cheyenne Munson, and Bob McMurray. Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach. *Journal of Phonetics*, 38:167–184, 2010.
13. Aniello de Santo and Thomas Graf. Structure sensitive tier projection: Applications and formal properties. Ms., Stony Brook University.
14. Samuel Eilenberg. *Automata, Languages, and Machines*. Academic Press, Inc., 1974.
15. G. Gallagher. Vowel height allophony and dorsal place contrasts in Cochabamba Quechua. *Phonetica*, 73:101–119, 2016.
16. E. Mark Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.
17. Thomas Graf. Comparing incomparable frameworks: A model theoretic approach to phonology. *University of Pennsylvania Working Papers in Linguistics*, 16, 2010.
18. Thomas Graf. The power of locality domains in phonology. *Phonology*, 34:385–405, 2017.
19. Kenneth Hale. Review of Hohepa (1967). *Journal of the Polynesian Society*, 77:83–99, 1968.
20. Jeffrey Heinz. Learning long-distance phonotactics. *Linguistic Inquiry*, 41:623–661, 2010.
21. Jeffrey Heinz. The computational nature of phonological generalizations. In *Phonological Typology*. 2018.
22. Jeffrey Heinz, Anna Kasprzik, and Timo Kötzing. Learning with lattice-structure hypothesis spaces. *Theoretical Computer Science*, 457:111–127, 2012.
23. Jeffrey Heinz, Chetan Rawal, and Herbert G. Tanner. Tier-based strictly local constraints for phonology. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Shortpapers*, pages 58–64, 2011.
24. Larry M. Hyman. Phonologization. In A. Julliand, editor, *Linguistic studies offered to Joseph Greenberg*, volume 2, pages 407–418. Anma Libri, Saratoga, 1976.

25. Adam Jardine. Computationally, tone is different. *Phonology*, 33:247–283, 2016.
26. Adam Jardine and Jeffrey Heinz. Learning tier-based strictly 2-local languages. *Transactions of the ACL*, 4:87–98, 2016.
27. Adam Jardine and Kevin McMullin. Efficient learning of tier-based strictly k-local languages. In F. Drewes, C. Martín-Vide, and B. Truthe, editors, *Language and Automata Theory and Applications, 11th International Conference*, pages 64–76. Springer, 2017.
28. C. Douglas Johnson. *Formal Aspects of Phonological Decription.* Mouton, The Hague, 1972.
29. Ronald Kaplan and Martin Kay. Regular models of phonological rule systems. *Computational Linguistics*, 20:331–378, 1994.
30. Vern M. Lindblad. *Neutralization in Uyghur.* University of Washington, 1990.
31. Sharon Y. Manuel. The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America*, 88:1286–1298, 1990.
32. Sharon Y. Manuel and Rena A. Krakow. Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research*, 77/78:69–78, 1984.
33. Connor Mayer and Travis Major. On the productivity of uyghur backness harmony, In progress.
34. Kevin McMullin. *Tier-based locality in long-distance phonotactics: Learnability and typology.* PhD thesis, University of British Columbia, 2016.
35. Robert McNaughton and Seymour Papert. *Counter-Free Automata.* MIT Press, 1971.
36. John Ohala. The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, and M. F. Miller, editors, *Papers from the parasession on language and behavior*, pages 178–203. Chicago Linguistic Society, Chicago, 1981.
37. ‘Ōiwi Parker Jones. Phonotactic probability and the Māori passive. *Proceedings of the Tenth Meeting of the ACL Special Interest Group on Computational Morphology and Phonology*, pages 39–48, 2008.
38. Jean Eric Pin. *Varieties Of Formal Languages.* Plenum Publishing Co., 1986.
39. Jose Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*, 2017. R package version 3.1-131.
40. R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2017.
41. James Rogers, Jeffrey Heinz, Margaret Fero, and Jeremy Hurst. Cognitive and sub-regular complexity. In Glyn Morrill and Mark-Jan Nederhof, editors, *Formal Grammar, volume 8036 of Lecture Notes in Computer Science*, pages 90–108. Springer, 2013.
42. James Rogers and Geoffrey K. Pullum. Aural pattern recognition experiments and the subregular hierarchy. *Journal of Logic, Language and Information*, 2011.
43. Robert Nathaniel Sanders. *Opacity and sound change in the Polish lexicon.* PhD thesis, UCSC, 2003.
44. M.P. Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8:190–194, 1965.
45. Imre Simon. Piecewise testable events. In H. Brakhage, editor, *Automata Theory and Formal Languages 2nd GI Conference*, volume 33 of *Lectures Notes in Computer Science*, pages 214–222, Berlin, 1975. Springer.
46. John Sylak-Glassman. The effects of post-velar consonants on vowels in Ditidaht. In Natalie Weber, Emily Sadlier-Brown, and Erin Guntly, editors, *Papers for the*

International Conference on Salish and Neighbouring Languages 49, University of British Columbia Working Papers in Linguistics, volume 37. 2014.

47. Bert Vaux. Disharmony and derived transparency in Uyghur vowel harmony. *Proceedings of NELS 30*, pages 671–698, 2000.
48. Anssi Yli-Jyrä. Describing syntax with star-free regular expressions. In *EACL*, 2003.
49. Anssi Mikael Yli-Jyrä et al. On finite-state tonology with autosegmental representations. In *Proceedings of the 11th International Conference on Finite State Methods and Natural Language Processing*. Association for Computational Linguistics, 2013.