Inducing Nonlocal Constraints From Baseline Phonotactics*

Maria Gouskova and Gillian Gallagher, New York University

November 14, 2018

Abstract

Nonlocal phonological patterns such as vowel harmony and long-distance consonant assimilation and dissimilation motivate representations that include only the interacting segments—projections. We present an implemented computational learner that induces projections based on phonotactic properties of a language that are observable without nonlocal representations. The learner builds on the base grammar induced by the Hayes and Wilson MaxEnt Phonotactic Learner (Hayes and Wilson 2008). Our model searches this baseline grammar for constraints that suggest nonlocal interactions, capitalizing on the observations that (a) nonlocal interactions can be seen in trigrams if the language has simple syllable structure, and (b) nonlocally interacting segments define a natural class. We show that this model finds nonlocal restrictions on laryngeal consonants in corpora of Quechua and Aymara, and vowel co-occurrence restrictions in Shona.

1 Introduction

Nonlocal phonological interactions such as vowel harmony and consonant dissimilation are a long-standing challenge for phonological theory. A key observation about such patterns is that the interacting segments define a natural class, and this is reflected in formal analyses either through feature geometric structures that constrain phonological patterns (Mester 1986; McCarthy 1988) or fixed scales of constraints that reflect natural class structure (Hansson 2001; Rose and Walker 2004). We present an inductive model that

*For helpful feedback, we would like to thank Arto Anttila and the anonymous reviewers of NLLT, as well as Maddie Gilbert, Juliet Stanton, Ildi Emese Szabó, Sora Heng Yin, Jon Rawski, audiences at OCP 2018 in London, UMass Amherst, Stony Brook, and the Phonology Winter School in Israel. Finally, we would like to thank Daniel Ridings for making the ALLEX corpus wordlist available to us, and Colin Wilson for sharing the code for the gain-based MaxEnt Phonotactic Learner, as well as detailed feedback on related work. This research was supported in part by NSF BCS-1724753 to the authors.

incorporates this insight about the role of natural class structure in nonlocal representations without assuming a predefined feature geometry or constraint set. Our learner attends to certain properties of a language that are observable without nonlocal representations, and searches for nonlocal constraints on projections defined by the natural class structure of the language. We demonstrate the success of our learner with three case studies, including co-occurrence restrictions on stops in Quechua (exemplified in (1)), the similar restrictions in Aymara, and vowel co-occurrence restrictions in Shona verbs (see (2)).

(1) Consonant co-occurrence restrictions in Quechua, in brief

a. initial ejectives and aspirates allowed:	k'utuj	'to cut'	k^{h}anij	'to bite'
b. medial ejectives and aspirates allowed:	rit'i	'snow'	jut^{h}u	'partridge'
c. no stop-ejective combinations:	*kut'u	*k'ut'u	*k^{h}ut'u	
d. no stop-aspirate combinations:	*kut^{h}u	*k'ut^{h}u	*k^{h}ut^{h}u	

(2) Vowel co-occurrence restrictions in Shona verbs, in brief

a.	[e e] but not [e i]:	-per-er-a	*-per-ir-a	'end in'
b.	[i i] but not [i e]:	-ip-ir-a	*-ip-er-a	'be evil for'
c.	[a i] but not [a e]:	-pofomaʣ-ir-a	*. . .aʣ-er-a	'blind for'
d.	[e u] allowed:	-svetuk-ir-a	*svetok-ir-a	'jump in'
e.	[o u] not allowed:	-pofomaʣ-ir-a	*pofu. . .	'blind for'

Our inductive learner builds on the Maximum Entropy (MaxEnt) Phonotactic Learner of Hayes and Wilson (2008). This learner works from positive learning evidence, in the form of the phonological words of the language, and searches through the space of possible n-gram constraints on natural classes to identify constraints that penalize underattested or unattested structures. While the Hayes and Wilson model is successful at finding phonologically meaningful local generalizations, this kind of learning is computationally intensive and does not scale up to searching through an exhaustive space of nonlocal interactions. Hayes and Wilson demonstrate that their learner can find nonlocal generalizations when supplied with projections by the analyst, but these generalizations cannot be captured without projections, and their model does not learn the projections on its own. We augment their model with a procedure that identifies nonlocal interactions and encodes them in projection-based constraints.

Our model is based on a key empirical insight about the local phonology of languages with nonlocal phonological interactions: while nonlocal restrictions hold at arbitrary distances, they may also be observable within a trigram. In many languages with nonlocal phonology, the interacting classes are frequently separated by only a single segment: in languages with consonant dissimilation and assimilation, the interacting consonants are often separated by just one vowel, $\underline{C}V\underline{C}$, and in languages with vowel harmony, there is often just one consonant between the assimilating vowels, $\underline{V}C\underline{V}$. Interactions across a single segment can be captured via trigram constraints in the baseline grammar—the grammar with no projections—and used as a clue that there is a more general nonlocal interaction in the language. Our model identifies relevant trigram constraints in the baseline grammar and builds natural-class based projections from them. By working with a statistical learner and a simple, natural-class based projection induction procedure, our model conducts a targeted and efficient search for nonlocal interactions and is less likely to confuse accidental and systematic gaps. We begin by presenting our learner in detail (§2) and then demonstrate how it works with three case studies (§3–5).

2 An Inductive Projection Learner

The baseline algorithm for our learner is the Hayes and Wilson MaxEnt Phonotactic Learner, described in § 2.1. This inductive learner is based on the principle of Maximum Entropy (Della Pietra et al. 1997; Goldwater and Johnson 2003; Hayes and Wilson 2008; Zuraw and Hayes 2017). The learner induces a grammar from learning data by searching through a space of possible constraints and evaluating these constraints for their usefulness in accounting for patterns in the learning data. The model selects a set of constraints and assigns these constraints weights, resulting in a grammar that assigns scores to novel forms. To this learner, we add a procedure for inducing projections on which nonlocal phonological interactions can be learned. Our model has two components, described in §2.2–2.3. First, the model evaluates the baseline grammar produced by the Hayes and Wilson MaxEnt Phonotactic Learner for evidence that a projection may be needed. Second, the model creates projections based on the output of the baseline grammar and builds a final grammar by searching these projections for useful constraints.

2.1 An overview of the MaxEnt Phonotactic Learner

The Hayes and Wilson MaxEnt Phonotactic Learner (Hayes and Wilson 2008) uses positive evidence (and implicit negative evidence) to induce phonotactic constraints against sequences that are unattested or underattested in a language. The learner is given a list of attested words and the features that describe the segments of the language.[1] The learner begins by constructing a list of natural classes and an exhaustive list of all possible *n*-gram constraints built from those natural classes. The learner then constructs its own list of hypothetical forms by combining the language's segments randomly, and uses an iterative scaling algorithm (Della Pietra et al. 1997) to identify unattested or underattested *n*-grams in the learning data. The learner induces *n*-gram constraints against the relevant sequences and uses the principle of Maximum Entropy to weight the constraints, maximizing the probability of the observed phonotactic distribution in the language. The output of the learner is a list of constraints and their weights, which can be used to assign probabilities and harmony scores to previously unseen data such as nonce words.

Constraint generation. The learner takes the phonological feature set defined by the analyst, identifies all the unique natural classes in it (using the shortest featural description of the class), and generates a space of all possible *n*-gram constraints (up to a certain *n*) composed of those natural classes. Phonological constraints can be paradigmatic (unigram) or syntagmatic (bigram, trigram, etc.). For example, Russian does not include a velar nasal at all, motivating a paradigmatic unigram constraint *[+dorsal, +nasal], whereas in English, velar nasals are prohibited in word-initial position, captured by the bigram *#[+dorsal,+nasal]. Accounting for the full range of phonological patterns requires constraints that span at least three positions—trigrams (see Goldsmith and Riggle 2012, inter alia). Trigrams are needed to capture phonological patterns such as intervocalic voicing (*VÇ̥V), or restrictions at word edges (e.g., *#[+dorsal, +nasal] in English). As Hayes and Wilson explain (2008:392), the number of possible natural class-based constraints grows exponentially with the size of the *n*-gram window, so it is in practice difficult to search even through a space of relatively short constraints when the natural classes exceed a certain number. The problem of distinguishing between systematic and accidental gaps also increases with the length of constraints, as discussed further in Wilson and Gallagher (2018).

In order for a constraint to be added to the grammar, it must meet or exceed the selection criterion (O/E or gain, discussed below). Since there are many constraints that may meet the criterion, Hayes and Wilson (2008) add several heuristics, inspired by phonological reasoning. These heuristics include a preference for shorter constraints, and a preference for constraints that mention larger natural classes over smaller ones.

[1] An anonymous reviewer asks how crucial it is to assume that the segmental inventory is given in advance. This is an interesting question, since traditional phonological reasoning about analyzing segmental inventories does usually depend on phonotactics: for example, the analysis of English [ʧ] as an affricate and [ts] as a cluster relies on distributional information. We do not attempt to solve this complex problem here, though see §5 for some related discusssion.

Constraint selection criterion. The original version of the learner distributed in 2008 uses the Observed/Expected (O/E) statistic to identify the most promising constraints (Trubetzkoy (1939, pp. 264-266). The O/E statistic calculates the likelihood of a sequence of X and Y given the independent probabilities of X and Y, allowing a distinction between phonologically meaningful underattestation and accidental gaps due to overall rarity of X or Y. The O/E statistic has been used extensively as a descriptive tool in work on probabilistic phonological constraints (Frisch et al., 2004; Gallagher and Coon, 2008; Coetzee and Pater, 2008), where the O/E calculation is position specific, with the relevant positions being defined by the analyst based on relevant phonological properties. The O/E metric in the 2008 learner, however, is not position specific, and Wilson and Obdeyn (2009) demonstrate that it is vulnerable to overestimating prohibitions when either X or Y is *positionally restricted.* This is an issue in our case studies: Quechua and Shona restrict some of the non-locally interacting segments sequentially and/or positionally, and the value of a nonlocal co-occurrence constraint needs to be assessed independently of these other restrictions. We therefore use an alternative heuristic for selecting constraints from the list of all possible n-grams, the *gain* criterion[2] (Della Pietra et al. 1997; Wilson and Gallagher 2018). The gain of a constraint is a function of the log likelihood of the model were the constraint to be added to the grammar without changing the weights of any of the constraints already in the grammar. Gain is set at a specific threshold; the higher the gain, the more statistical support is required for a constraint to be added to the grammar.

2.2 Exploring the baseline grammar for placeholder trigrams

The Hayes and Wilson MaxEnt Phonotactic Learner augments its list of natural classes by a [word boundary] feature, to track phonological effects at word edges. Word edges are [+word boundary] (+wb), and [-word boundary] refers to all of the consonants and vowels in the language. We refer to [-word boundary] (henceforth [-wb] or simply [ ]) as a *placeholder class.* Since the the placeholder class is the largest natural class in any language, the learner's bias towards large natural classes will make it likely to refer

[2]Della Pietra et al. (1997, 4) characterize gain as "the improvement [a constraint] brings to the model when it has weight [w]":$Gain_{Con}(w, C) = D(\tilde{p}||Con) - D(\tilde{p}||Con_{wC})$, where C is the constraint with the weight w, D is the Kullback-Leibler divergence, $\tilde{p}$ is the probability distribution of the data, and *Con* is the current constraint grammar.

Della Pietra et al. explain the reason for this method of calculating gain intuitively as follows: "We approximate the improvement due to adding a single candidate [constraint], measured by the reduction in Kullback-Leibler divergence, by adjusting only the weight of the [constraint] and keeping all of the other parameters of the [grammar] fixed. In general this is only an estimate, since it may well be that adding a [constraint] will require significant adjustments to all of the parameters in the new model. From a computational perspective, approximating the improvement in this way can enable the simultaneous evaluation of thousands of candidate [constraints], and makes the algorithm practical." (We modified the language slightly to translate it into constraint/grammar terms.) We might add that defined in this way, gain can be calculated for each constraint even when the grammar contains no constraints yet, whereas for O/E, there needs to be an arbitrarily set threshold.

to the placeholder whenever the generalization is consistent with the data. For example, take a strict CV language in which [k] and [q] never occur across an intervening vowel. A linguist might state this generalization as *[k]V[q], but the Hayes and Wilson MaxEnt Phonotactic Learner would induce the more general constraint *[k][ ][q], since neither vowels nor consonants occur in the medial position.

The intuition behind our projection induction procedure is that trigram constraints with the placeholder class as the medial gram are a cue to the learner that classes on either side interact nonlocally. A constraint *X[-wb]Y tells the learner that X and Y interact phonologically, and that the identity of the segment between them is irrelevant—this is precisely the characteristic of a nonlocal phonological interaction. We take the presence of such constraints in the baseline grammar to indicate the need to explore nonlocal co-occurrence restrictions between X and Y by looking for generalizations that hold on projections defined by natural classes that include both X and Y.

Segmental trigram constraints with a placeholder segment often capture a piece of a nonlocal interaction, but the whole interaction cannot be captured without a projection. In Quechua, for example, the restriction on stops followed by ejectives is partially accounted for by the trigram constraint *[-continuant, -sonorant][-wb][+constricted glottis] (henceforth [cg]) on the baseline projection, which penalizes unattested forms like *[kap'i], with one segment intervening between the stop and the ejective. But stops also cannot be followed by ejectives when more segments intervene, as in *[kasp'i] or *[kamip'a]. To account for the full pattern, a projection with only oral stops is needed. Similarly, in Shona, interactions between vowels are partially captured on the baseline projection with a trigram constraint *[-high, -back][-wb][-high, -low, +back]. This constraint bans certain vowels separated by a single consonant, e.g., *[epo], but interactions between vowels that are separated by more than one consonant require a projection that includes only vowels, e.g. *[empo].

The success of this induction strategy depends on the syllable structure of the language and on the positional distribution of segments. The learner will notice co-occurrence restrictions on consonants when they are frequently separated by just one vowel, CVC, and vowel restrictions are easiest to notice when the vowels are usually separated by just one consonant, VCV. As we will show in our case studies, this is true of Quechua, Aymara and Shona—even though all languages tolerate deviations from strict CV alternation, the CVC and VCV configurations are frequent enough in the learning data that trigram constraints with a placeholder class are reliably included in the baseline grammar. The observation that predictable syllable structure makes non-local relations easier to detect suggests a plausible learning-based explanation for

McCarthy's (1989) hypothesis that templaticism leads to planar segregation of consonants and vowels. C-to-C and V-to-V interactions will be most noticeable to the learner in languages with the simplest or most predictable syllable structure, since the learner can see these interactions in segmental trigrams.

On the other hand, languages with complex syllable structure may not show the segments from classes X and Y in trigram configurations sufficiently often for the learner to notice a co-occurrence restriction. In a language with more complex syllable structure—such as Russian—any dependencies between noncontiguous vowels or consonants would be much harder for the learner to detect. Even in such languages, CVC trigrams are more common than CCC and so on, but relatively frequent VCV and CVC trigrams alone is no guarantee that all nonlocal interactions will be observable in trigrams.[3] For example, in a language where [l] and [r] dissimilate, it is not sufficient that there be many CVC strings; rather, there must be sufficiently many liquid-V-liquid strings for the learner to notice the rhotic/liquid combinatorics in particular (as opposed to accounting for the underattestation of liquid-V-liquid strings through bigram constraints). We return to this aspect of the learning data throughout the paper.

2.3 Creating non-baseline projections

After identifying placeholder trigram constraints on the baseline projection, the learner constructs a nonlocal projection (e.g., a projection including only oral stops, or only vowels) for each constraint and builds a final grammar by searching through the baseline projection and all nonlocal projections for constraints.

Bigrams and trigrams only. While the learner searches for unigram constraints on the baseline projection, we do not allow it to posit such constraints on other projections. At the baseline level, a unigram constraint can indeed be a reasonable way to capture the phonotactics of a language—e.g., [ʒ] is relatively rare in English, so its distribution may be well captured with a unigram constraint. On higher projections, however, unigram constraints are nonsensical. Non-baseline projections are postulated to capture *interactions* between non-adjacent segments, so we restrict the search space on these projections to bigram and trigram constraints.

Which classes define a projection. When the learner identifies a placeholder trigram *X[-wb]Y, it constructs a projection from the smallest natural class that contains all the segments in both X and Y. Very

[3] We did the counts for a transcribed Russian dictionary of 103,000 words. Looking at consonants in trigram and tetragram configurations, $\underline{C}V\underline{C}$ accounted for 337,415 or 63% of all the combinations; $\underline{C}C\underline{C}$: 18,516 (3.4%), $\underline{C}CV\underline{C}$: 76,574 (14%), $\underline{C}VC\underline{C}$: 93,637 (18%), $\underline{C}VV\underline{C}$: 7,946 (1%). For vowel-to-vowel n-grams, the counts are $\underline{V}C\underline{V}$: 117,214 (64%), $\underline{V}CC\underline{V}$: 61,344 (33%), $\underline{V}CV\underline{V}$: 2,074 (1%), $\underline{V}VC\underline{V}$: 2512 (1%). We give comparable numbers for other languages, where relevant, in their respective sections.

often, this is either X or Y itself: e.g., *[-sonorant, -continuant][-wb][+cg] will give rise to a [-sonorant, -continuant] projection, since ejectives are a subset of plosives. If neither class is a superset of the other, then the smallest class that is a superset of X and Y will be searched.

A projection based on the smallest natural class that includes both X and Y represents the maximally general hypothesis that all intervening segments that do not belong to the class are irrelevant. This will be the correct hypothesis provided (i) the baseline grammar includes the most general placeholder trigram constraint that accounts for the restriction, and (ii) the interaction does not involve segments outside of the class, i.e., no special class Z is transparent or opaque to the interaction between X and Y. We elaborate on both of these points in §5.5 and §6.3 below.

Which features are visible on the projection. Our learner considers the full space of features on all projections. Any non-zero feature in the natural class defined by the projection is visible—this includes features with $\pm$ values and privative features that have + values only. We see this choice as representing the null hypothesis. Following Hayes and Wilson (2008), [$\pm$wb] is always projected as well; this is necessary to encode positional and ordering generalizations (e.g., in Shona, [o] and [e] are never the last vowels in a verb stem, so *[-high, -low][+wb] is a sensible bigram on the vowel projection).

2.4 Why not search exhaustively?

The Hayes and Wilson MaxEnt Phonotactic Learner's ability to find placeholder constraints opens up a logical possibility: suppose that, instead of constructing a projection for a superset natural class containing X and Y from constraints against [X][ ][Y] trigrams, we instead allow the learner to consider trigram constraints that ignore an arbitrarily long string of interveners between X and Y: [X][ ]*[Y], trigrams with zero to any number of placeholders. This would allow the learner to capture nonlocal interactions at arbitrary distances without including nonlocal projections. We argue in this section that this alternative is not viable for real language data.

Algorithms are evaluated by how they scale up with the size of the problem, so we must consider the combinatorics of *n*-gram searches. In our learner, all words are decomposed at most into local trigrams. On the baseline, these are strings of three adjacent segments or of the natural classes to which they belong (e.g., [patu] contains {p,a,t}, {a,t,u}, but {p,a,t} also expands to the trigrams [+cons][-cons][+cons], [-voice,-cont][+low,+back][-voice,-cont], and so on). Since nonlocal projections are defined by natural class membership, the number of projection-based trigrams is always smaller than the number of segments in the word; thus, the [+syllabic] projection representation for [patu] includes only a bigram [a u], which expands to [-high,+low,+back,-round][+high,-low,+back,+round], or [+low][-low], or [-round][+round], and so on. The number of natural class n-grams, as opposed to segmental n-grams, depends on the segmental inventory of the language and the features assumed. Even without considering the relationship between the number of segments and the number of natural classes they belong to, however, it is easy to demonstrate that the number of *nonlocal* segmental *n*-grams dwarfs the number of *local* ones.

The number of segmental *n*-grams in a word is a linear function of the length of the word, as shown in (3). In the formula below, *l* is the number of segments in the word, and *n* is the length of the *n*-gram window. (We sidestep the fact that edges of words must be treated as trigrams as well, as in #pr, pa#, since the two extra *n*-grams do not make much of a difference for this comparison.)

(3) The number of local segmental *n*-grams in a word

$$N_{ngrams} = l - n + 1$$

On the other hand, the number of nonlocal ordered substrings (length *n*) of a word (length *l*) is calculated as a product of factorials:

(4) The number of nonlocal segmental *n*-grams in a word

$$N_{ngrams} = l!/n!/(l-n)!$$

Example (5) illustrates the number of local and nonlocal segmental trigrams contained in words of lengths from 3 to 5. The number of trigrams that include the edgemost segments is one for local calculation, but it grows fast with the length of the word for nonlocal calculations:

(5) Local and nonlocal trigrams in words of 3, 4, and 5 segments

	Local trigrams	N(local)	Nonlocal trigrams	N(nonlocal)
pat	pat	1	pat	1
patu	pat, atu	2	pat, atu, pau, ptu	4
patuk	pat, atu, tuk	3	pat, atu, tuk, pau, pak, puk, ptu, ptk, atk, auk	10

The problem is exacerbated when we look at trigrams of natural classes rather than segments. The number of natural classes each segment belongs to varies with the segment—so the number of natural class trigrams is a product of the natural classes each segment belongs to, i.e., for [pat], *n* is not 3 but rather

$C_p \times C_a \times C_t$, where C_p is the number of natural classes C containing [p]. Languages vary in the number of natural classes—it depends on the number of segmental contrasts (and analytically, on the feature system assumed); languages also vary in the length of words, and the complexity of segmental phonotactics. The nature of the learning data can affect the success of various phonological learning algorithms dramatically (Stanton to appear and others). To examine the combinatorics of natural class trigrams, we must look at some real language corpora.

We counted local and nonlocal trigrams in six natural languages: Aymara, Quechua, Shona, Hungarian, Mongolian, and Russian. The first three languages are our case studies; Hungarian and Mongolian both have vowel harmony (Siptár and Törkenczy 2000; Svantesson et al. 2005); Russian is included as an example of a language with a large number of natural classes.[4] The number of natural classes was computed by the Hayes and Wilson MaxEnt Phonotactic Learner. We calculated the number of natural class trigrams for each language based on real word corpora (included in the supplemental materials, along with the features we used); the relevant quantitative properties of these corpora are summarized in Table 1. For example, in Quechua, there are 203 natural classes. As shown in Table 2, an average word in our Quechua corpus contains 128,377 local natural class trigrams: thus, [patu] contains two local segmental trigrams {p,a,t}, {a,t,u}. These expand to multiple natural class trigrams: {p,a,t} can be rewritten as [-cont][+son][+cor], [-voice][+syll][-cont], etc. If trigrams are nonlocal (that is, [patu] = {p,a,t}, {a,t,u}, {p,t,u}, {p,a,u}), then an average word has 2,316,929 natural class-based trigrams.

	Number of nat. classes	Mean wd length (in segments)	Wds in corpus
Aymara	89	4.54	1,960
Quechua	203	9.37	10,843
Shona	236	9.57	300,443
Hungarian	394	8.31	71,136
Mongolian	275	6.85	48,942
Russian	406	9.27	86,836

Table 1: Summary statistics over six natural language corpora

[4] Padgett (1991) does report a gradient co-occurrence restriction in 500 Russian roots; see also Kochetov and Radisic (2009).

	Local trigrams/wd			Nonlocal trigrams/wd		
	Mean	Min	Max	Mean	Min	Max
Aymara	4,222	660	13,928	13,440	891	144,018
Quechua	128,377	5,508	402,254	2,316,929	5,508	24,357,380
Shona	78,213	1,728	338,124	1,364,777	1,728	23,056,995
Hungarian	198,457	10,773	731,202	2,878,029	10,773	48,498,150
Mongolian	171,744	14,950	630,437	1,597,763	14,950	35,699,484
Russian	984,428	34,596	3,402,677	18,378,498	34,596	283,814,090

Table 2: Natural class-based trigrams per word in six languages, calculated over local and nonlocal substrings. The mean, minimum and maximum number of local and nonlocal trigrams per word are shown.

The difference for Quechua is about 18 times more trigrams to consider per word, on average. An 18-fold difference might not seem like much, but as shown in Figure 1, the differences quickly add up as words get longer. Since words get quite large in agglutinative and highly inflecting languages, this is a serious concern. Note the scale of the plot: the y-axis runs to 45,000,000 trigrams *per word*. For a corpus such as Aymara roots, the difference between local and nonlocal trigram calculations is negligible. For word corpora of languages with complex segmental inventories, such as those of Hungarian, Mongolian, or Russian, the numbers diverge dramatically.

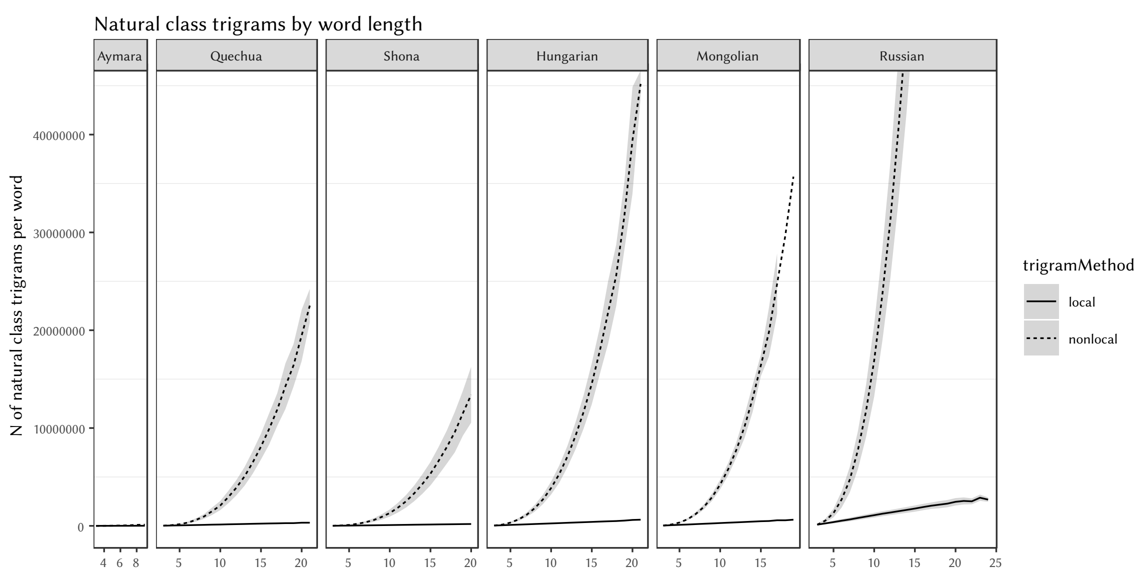

Figure 1: Average (mean) number of natural class-based trigrams per word, as a function of word length (in segments) in six natural language corpora. Gray areas show standard deviation of the mean. Solid lines show numbers reckoned over segmentally local strings; dashed lines—nonadjacent linearly ordered segments.

Devising a computationally efficient search through nonlocal trigrams composed of natural class matrices will require a sophisticated implementation that to the best of our knowledge is currently lacking. Even in our method, some languages push the computational learner to its limits (see §6.2, as well as Hayes and Wilson 2008:392). Our proposal implements a targeted search for nonlocal interactions, based on properties of a language that are observable from a local *n*-gram model. In addition to avoiding the considerable computational challenge of an exhaustive search, our method zeroes in on classes that are known (from the baseline grammar) to interact nonlocally, and thus also limits the likelihood that the grammar will stumble onto accidental gaps. We now turn to illustrating our model with three case studies.

3 Quechua

To illustrate the basic insight and procedure of our learner, we begin with the case study of categorical laryngeal phonotactics in South Bolivian Quechua (henceforth just "Quechua"). We show that the baseline grammar for Quechua includes trigram constraints that capture pieces of the co-occurrence restrictions in the language, and that the projection induced from these constraints results in a grammar that distinguishes legal from illegal nonce forms via concise, highly-weighted constraints on the nonlocal projection.

3.1 Laryngeal restrictions in Quechua

Quechua contrasts three series of stops: plain (voiceless unaspirated) [p t ʧ k q], ejective [p' t' ʧ' k' q'] and aspirate [pʰ tʰ ʧʰ kʰ qʰ]. Affricates pattern with stops both in terms of laryngeal contrasts and in phonotactic distribution. Stops are subject to numerous distributional restrictions:

(6) Restrictions on stops in Quechua

a. Roots contain ejectives, aspirates, and plain stops; suffixes can only have plain stops (e.g., *-ŋk'u, ✓-ŋku).

b. Stops are only permitted in onset position; codas must be fricative or sonorant consonants (*map.ta, ✓man.ta, ✓mas.k'a).

c. Ejectives and aspirates can only occur non-initially if preceded by fricatives or sonorant consonants.[5]

[5] Aspirates may also appear in vowel-initial words, though ejectives are absent from such forms. See Gallagher (2015) for discussion.

The combinatorical restrictions on ejectives and aspirates are our focus, and we illustrate them in more detail in Table 3. As shown in the table, ejectives and aspirates may occur initially in a root, or in medial position in roots where the initial consonant is not a stop (a fricative or sonorant). Ejectives and aspirates may not occur in medial position in roots that have a plain, ejective or aspirate stop initially.

Attested combinations				Impossible combinations		
(a) tʃ'uspi	'fly'	(c) rit'i	'snow'	(e) *kup'i	(g) *k'up'i	(i) *k^{h}up'i
(b) k^{h}utʃi	'pig'	(d) ʎimp^{h}u	'clean'	(f) *kup^{h}i	(h) *k'up^{h}i	(j) *k^{h}up^{h}i

Table 3: Quechua laryngeal restrictions

Quechua speakers' sensitivity to these restrictions has been demonstrated in a variety of behavioral experiments (Gallagher 2015, 2016), which find effects in production, perception and nonce word acceptability judgments for all unattested stop combinations. The restrictions on stops in Quechua can be grouped under just two generalizations about sequences of nonadjacent natural classes: *[-cont, -son]...[+constricted glottis] and *[-cont, -son]...[-cont, +spread glottis] (note that aspirates must be picked out as [-continuant, +sg] to distinguish them from [h], which is also [+sg]). While the restrictions are typically described as being restrictions on roots, the absence of ejectives and aspirates from affixes in the language means that the restrictions hold categorically at the word level as well.

In addition to the restrictions on combinations of stops, Quechua consonants show other distributional gaps that we do not explore in great detail here. First, aspirates are absent from roots with initial [h], though Gallagher (2015) shows that the psychological reality of this restriction for Quechua speakers is questionable. Second, uvulars [q q' q^{h}] and velars [k k' k^{h}] do not co-occur within roots, though they may co-occur across morpheme boundaries within a word; this restriction is explored in Wilson and Gallagher (2018).

3.2 Methods: the training and testing data

We trained our model on a corpus of 10,848 phonological words (available at on GitHub (github.com/gouskova/inductive_pr compiled from 31 issues of the Bolivian Quechua newspaper *Conosur Ñawpaqman*, published by CENDA and available at http://www.cenda.org/periodico-conosur.[6] The word corpus was manually checked to re-

[6]While the newspaper is primarily a Quechua language periodical, it includes numerous articles in Spanish, as well as Spanish phrases and Spanish roots embedded in Quechua text. The majority of Spanish forms were removed from the word corpus, including Spanish words that were inflected with Quechua morphology. The only exception to this are those words, mostly place names, that are consistent with the native phonotactics of Quechua.

move Spanish and corrected for mispellings. The orthographic corpus was then phonetically transcribed. The phonetic transcription represented nasal place assimilation (of coda nasals to following obstruents), as well as vowel lowering by a uvular consonant and retraction of coda consonants preceding a uvular. Laryngeal distinctions in Quechua are represented with two privative features, [+constricted glottis] ([+cg]) for ejectives and [+spread glottis] ([+sg]) for aspirates.

To test the grammars that the model learns, we created a large set of phonotactically legal and illegal nonce forms. The nonce forms were all disyllabic (C)VCV, (C)VCCV—the canonical root shapes in the language. While the testing sets were large, they were not exhaustive, and were designed to test specifically whether the models capture the distribution of stops in the language. The testing words were all (C)V(C)CV forms that (i) included at least one stop, (ii) respected nasal assimilation and uvular retraction and, (iii) only included CC clusters that were attested in the training corpus. Forms with a single stop are all classified as 'legal', as are forms with an initial stop and a medial plain stop; forms with an initial stop (plain, ejective or aspirate) and a medial ejective are classified as 'illegal-ejective', and those with an initial plain stop and medial aspirate are classified as 'illegal-aspirate'. The testing set included 24,352 forms (18,502 legal, 3,645 illegal-ejective and 2,205 illegal-aspirate).

3.3 The baseline grammar

We first look at the output of the baseline grammar—the grammar with no projections—to see whether it includes placeholder trigram constraints that capture part of the nonlocal phonotactics of the language. If the laryngeal restrictions can be detected as an underattested trigram in this model, we expect the baseline grammar to include the constraints *[-sonorant, -continuant][ ][+cg] and *[-sonorant, -continuant][ ][-continuant, +sg]. These constraints penalize illegal forms such as *[p'ak'a], but they are not violated by forms where the illegal combination of consonants is separated by more than a single segment, e.g., *[p'ask'a].

Which constraints make it into the baseline grammar depends on the minimum gain threshold supplied by the analyst and the amount of training data. Generally, the lower the gain, the more likely it is that a given constraint will be learned, but smaller data sets also require lower gain than larger data sets.[7] The

[7] Another parameter is whether the model is asked to look for violable or inviolable constraints. In either condition, whether a constraint is included in the grammar depends on its gain, but an inviolable constraint simulation only considers constraints whose observed violations are zero. To keep the amount of information digestible, we only consider inviolable constraint models of Quechua and Aymara, since the laryngeal phonotactics are categorical. The results reported here are replicable with similar settings for violable constraint models as well. For all models reported throughout this paper, we ran the model with a large

gain that most accurately represents the threshold that human learners use is an empirical question, to be tested by assessing the psychological reality of the generalizations captured by grammars with different gain levels. In our simulations, the baseline learner finds the laryngeal placeholder trigram constraints at both high and low gain. With gain set at 25, there are more than 200 constraints in the grammar, while only 20 constraints are included in a model with gain set at 200. Importantly, models at all gains include the target placeholder trigram constraints. We report models with 150 gain, representing a fairly conservative estimate of the constraints that a human Quechua learner may have in their grammar. The fit of a baseline model to the testing words is shown in Figure 2, which is a violin plot—a vertical density plot with dots showing the means. When considering CVCV, VCV, and VCCV forms (grouped together as "other" in the plot), the model distinguishes legal from illegal laryngeal combinations. Legal nonce words have harmony clustering around zero, whereas illegal ones have more violations—notably, violations of the trigram constraints—and therefore lower harmony. No distinctions are made between illegal and legal CVCCV forms, however, because the interacting consonants are separated by more than a single segment. All of those forms have high harmony scores (low constraint violations), regardless of actual phonotactic legality.

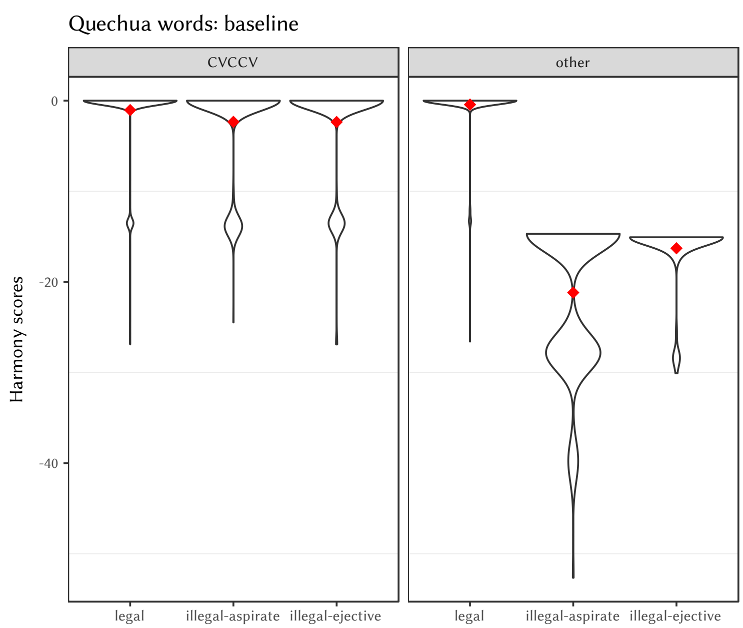

Figure 2: Quechua: harmony scores for nonce words, baseline grammar

enough constraint set that the model returned fewer constraints than it was asked for. This means that constraint set size was not an analyst-manipulated parameter that affected the fit of the model.

In this and following simulations, we evaluate both differences in the mean scores assigned to testing forms in a particular category, as well as separation between legal and illegal categories. We take a mostly holistic, qualitative approach to evaluating the grammars for two reasons. First, because we don't have a full set of behavioral results showing how speakers treat each of the forms in our large testing sets, we lack a detailed set of data that our grammars could be evaluated against. Second, since the grammars are learned directly from the statistical properties of the input data and contain many constraints, they often include several constraints that a linguist would be unlikely to posit. Such constraints account for the "tails" seen in this and other figures, where small numbers of forms in a given category are given particularly bad scores. For example, the baseline grammar for Quechua assigns a large penalty to the form [ɲerqʰa], which we classify as legal because it doesn't violate any known phonotactic constraints. The model, however, includes the constraints *[+sonorant, -anterior][+RTR, -back] (penalizing sequences of [ɲ ʎ j][e]) and *[-back][-syllabic, +continuant][+sg] (penalizing sequences of [e i][s ʃ x r l ʎ j w][pʰ tʰ ʧʰ kʰ qʰ]), which penalize [ɲerqʰa]. It is matter for future, empirical testing to determine whether constraints of this type, and the 'tails' that ensue, are examples of the model overfitting, or whether they represent true grammatical constraints (for some relevant work on inductive learning of phonotactics in English, see Hayes and White 2013). In designing our testing sets, we attempted to minimize the interference of orthogonal phonotactic restrictions, but our ability to do this was limited by available descriptions of the phonotactics of the languages in question.

The baseline model reliably finds the placeholder trigrams because of two properties of Quechua. First, all three laryngeal categories of stops appear with sufficient frequency in both initial and medial position within the word, as shown in Table 4. The absence of certain combinations stands out; it cannot be reduced to a local bigram constraint. This is in contrast to many languages with ejectives and aspirates, where these sounds are either restricted to absolute word initial position or are very rare outside of initial position (MacEachern 1997; Beckman 1998). In such a language, the absence of sequences such as [pak'...] can be captured by a bigram constraint on non-initial ejectives or aspirates (e.g., *[-wb][+cg]), and consequently there is no need in the model for longer trigram constraints. Indeed, such languages are reasonably described as not having nonlocal combinatorical phonotactics at all. While the proportions in Table 4 show that aspirates are generally less frequent than ejectives, and that both ejectives and aspirates are much less frequent than plain stops outside of initial position, they do still have a non-trivial frequency in non-initial position.

	plain	ejective	aspirate
initial	34%	8%	8%
non-initial	47%	2%	1%

Table 4: Quechua: frequency of plain, ejective and aspirate stops in initial and non-initial position. Percentages are out of all consonants, e.g., 34% of initial consonants are plain stops.

Second, the positions where the restricted segments occur in Quechua—onsets—are frequently separated by only a single vowel, as shown in Table 5. Under these conditions, the absence of stop-[ ]-ejective and stop-[ ]-aspirate combinations requires a trigram constraint. If Quechua were such that all or almost all syllables had coda consonants, stop-[ ]-ejective and stop-[ ]-aspirate combinations would still be unattested, but their absence would be attributable to a local bigram constraint against ejectives and aspirates in coda position (since C2 in a C1VC2 configuration would always or often be a coda consonant).[8]

Onset...onset n-grams	N of sequences	Proportion
$\underline{\text{C}}\text{V}\underline{\text{C}}\text{V}$	19,237	67%
$\underline{\text{C}}\text{VC}\underline{\text{C}}\text{V}$	9,310	33%

Table 5: Quechua: Onset-V-onset trigrams as percentage of all onset...onset pairs (sequences in 10,848 words were counted).

In sum, the distribution of natural classes and the frequency of syllable structures in Quechua allows nonlocal restrictions on combinations of stops to be reflected in a baseline grammar as placeholder trigram constraints.

3.4 Inducing projections and learning a final grammar

The baseline grammar includes the placeholder trigram constraints *[-sonorant, -continuant][ ][+cg] and *[-sonorant -continuant][ ][+sg, -continuant], which motivate a search through the [-sonorant, -continuant] projection. Given the feature specifications for Quechua, [-sonorant, -continuant] (the class of all oral stops) is the smallest natural class which includes both natural classes mentioned in the placeholder trigram; [+cg] segments and [+sg, -continuant] segments are subsets of the [-sonorant, -continuant] class. When the [-sonorant, -continuant] projection is included along with the baseline segmental projection, the model learns a final grammar that includes two general constraints that capture the full range of unattested

[8]A baseline grammar run on a modified Quechua training set where codas were added to all syllables confirmed that this is true; the grammar includes a highly weighted constraint against stop-consonant bigrams, but no trigram constraints on stop-[ ]-ejective or stop-[ ]-aspirate sequences.

stop combinations in the language: *[-wb][+cg] and *[-wb][+sg]. These constraints state that, when looking only at oral stops, ejectives and aspirates are always first in the word; that is, ejectives and aspirates are the leftmost stop in the word. This is the correct generalization:

(7) Projection-based representations for legal and illegal Quechua nonce words

[-cont, -son]	p	p'	p' t	p^h t'	t k^h
baseline (all segs)	p a m a	m a p' a	p' a t a	*p^h a n t' a	*t a s k^ha

These constraints are found and given high weights (weights of 15-16; compare the lower weights of some constraints in the Shona grammar in Table 5 below) in grammars with the full range of gains tested, and the resulting grammars show a separation between the scores assigned to legal and illegal nonce forms in the large testing set. To illustrate, we show the distribution of scores assigned to testing forms in a model with 150 gain in Figure 3. Unlike the baseline grammar shown in Figure 2 above, the final grammar with the [-continuant, -sonorant] projection distinguishes legal from illegal forms for both CVCCV nonce words and VC(C)V and CVCV ("other") nonce words; legal nonce words have few if any violations, and therefore harmony close to 0.

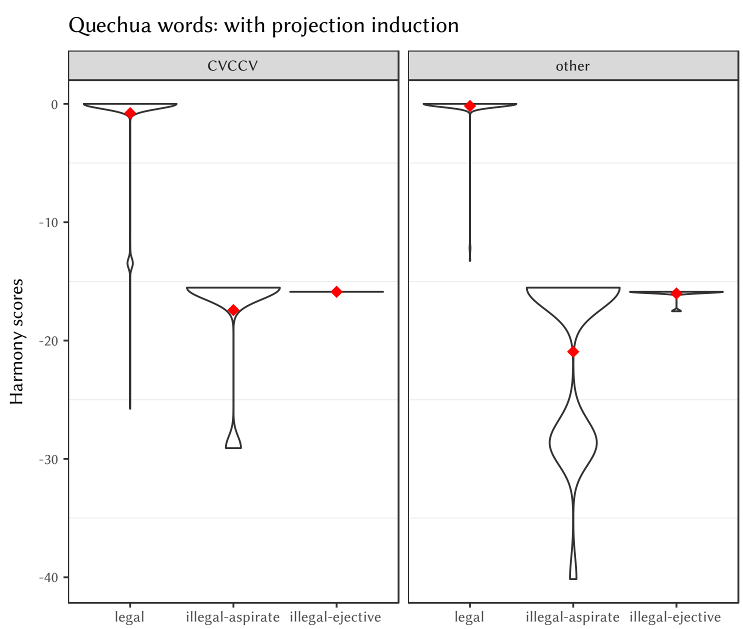

Figure 3: Quechua: harmony scores for nonce words, simulation with projection induction

Models with higher gain tend to include fewer constraints on underattested structures that phonologists would consider to be likely accidental gaps, as higher gain means that more statistical support is needed for a constraint to be added to the grammar. As can be seen in the figure, however, the Quechua model with high gain still includes constraints that penalize forms that we marked as legal—for example, a constraint against front vowel-continuant-aspirate sequences, like [ʎ**isk**^{h}u]. The 'tails' for each category in the figure are due to forms that violate trigram constraints of this sort. As discussed above, whether these specific trigram constraints represent real phonotactic restrictions in the language or are examples of the model overfitting and learning constraints on accidental gaps is an empirical question that we do not attempt to answer here.

3.5 Summary

The Quechua case study illustrates that nonlocal restrictions can be detected by examining local trigrams. This empirical observation offers a simple way to narrow down the search space of possible nonlocal interactions. Our learner examines the baseline grammar for placeholder trigram constraints of the form X[ ]Y and induces a nonlocal projection from these constraints based on the smallest natural class that includes both X and Y. In the next section, we demonstrate that the procedure can be generalized to the laryngeal restrictions in Aymara, before turning to the somewhat different case of Shona vowel harmony in §5.

4 Aymara

The Bolivian variety of Aymara is similar to Quechua both in the structure of roots and the laryngeal restrictions, though there are interesting differences (MacEachern 1997; Hardman 2001). The languages are not genetically related, though they are in contact with one another. As in Quechua, Aymara roots are primarily disyllabic (C)V(C)CV, with ejectives and aspirates occuring in onset position. Here we show that the baseline model for Aymara reliably includes multiple placeholder trigrams on unattested onset combinations, which motivate two nonlocal projections on which the full extent of the nonlocal restrictions can be captured.

4.1 Laryngeal restrictions in Aymara

Like Quechua, Aymara contrasts three series of stops, plain (voiceless unaspirated) [p t ʧ k q], ejective [p' t' ʧ' k' q'] and aspirate [pʰ tʰ ʧʰ kʰ qʰ]. The stops are subject to several combinatorial restrictions, summarized in Table 6. The phonotactics of Aymara are more permissive than in Quechua. As in Quechua, ejectives and aspirates may not follow plain stops in the root, and heterorganic ejectives may not co-occur in pairs. Unlike in Quechua, pairs of aspirates are permitted (heterorganic or identical), as are ejective-aspirate combinations and combinations of identical ejectives. Examples are from De Lucca (1987).

combination	example		Aymara	Quechua
initial ejective	k'awna	'egg'	✓	✓
initial aspirate	ʧʰiwi	'to sing'	✓	✓
fric/son-ejective	heq'e	'to smell'	✓	✓
fric/son-aspirate	laqʰa	'darkness'	✓	✓
identical aspirates	kʰuskʰa	'together'	✓	*
identical ejectives	t'ant'a	'bread'	✓	*
aspirate-ejective	pʰiʧ'u	'triangular'	✓	*
ejective-aspirate	k'utʰi	'thumb'	✓	*
aspirate-aspirate	pʰiʧʰa	'fire'	✓	*
plain-ejective	*piʧ'u	—	*	*
ejective-ejective	*p'iʧ'u	—	*	*
plain-aspirate	*piʧʰa	—	*	*

Table 6: Aymara laryngeal restrictions, with schematic comparison to Quechua.

The three restricted combinations in Aymara require three separate constraints *[+plain]...[+cg], *[+plain]...[+sg], and *[+cg]...[+cg]. Here, we assume a feature system with three privative features designating each of the three laryngeal classes. An alternative would be to use just two binary laryngeal features, with plain stops being picked out as [-cg, -sg]. The heuristics of Hayes and Wilson's learner make the model less likely to learn constraints on classes that require more features to pick out (recall §2.1), and so we opt for the privative feature option in order to put the three laryngeal classes on even footing with respect to the particularities of the baseline learner.[9]

[9] Indeed, a model where binary [sg] and [cg] are used does not include any constraints on plain stops. This could be interpreted as a failing of the heuristic in the Hayes and Wilson model, or it could be taken as evidence that privative features are a better hypothesis in this particular case.

4.2 Methods: the training and testing data

We tested our model on 1984 Aymara roots, extracted from De Lucca (1987). We used a root corpus instead of a word corpus because suffixes in Aymara may include ejectives and aspirates, introducing exceptions to the restrictions at the word level.[10] Our transcription represented vowel retraction of a uvular (which is represented in the Aymara orthography) and nasal place assimilation (which is not). To test the grammar that our model learns, we created a large set of phonotactically legal and illegal nonce forms, as for Quechua. The nonce forms were all disyllabic (C)VCV, (C)VCCV strings that contained at least one stop, included only consonant clusters attested in the training data and respected nasal assimilation and uvular retraction. Forms were classified as 'legal' or 'illegal' based on their status in Table 6 above. The testing set included 23,548 forms (23,548 legal, 1,389 plain-ejective, 1,108 plain-aspirate, 903 ejective-ejective).

As mentioned above, the laryngeal classes were represented with three privative features, [plain], [cg] and [sg]. The legality of identical pairs of ejectives in the language—what we will call the *identity exemption*—poses a representational challenge, both for Hayes & Wilson's learner and for other phonological models. The identity exemption can be captured for ejectives by treating them as a single segment (autosegmental spreading, e.g., MacEachern 1997; McCarthy 1989) or as standing in a correspondence relationship similar to reduplicated strings (Gafos 1999; Zuraw 2002; Rose and Walker 2004). Within inductive constraint models, the identity exemption to phonotactic restrictions has been accounted for by representing one of two identical consonants as a copy, using a placeholder segment X in the transcription (Colavin et al. 2010; Gallagher 2014). Under this method, a form like [t'ant'a] 'bread' is transcribed as [t'anXa], where 'X' is a segment bearing a single feature [+copy], as opposed to the full set of features that designate [t']. This representational choice, which we adopt, allows the model to find a constraint against non-identical ejectives, but it is orthogonal to the investigation of whether projections can be identified based on trigram constraints. Since the Hayes and Wilson learner has not yet been augmented with the capacity to represent algebraic constraints that explicitly reference matching or mismatching (though see Berent et al. (2012) for a proposal)[11], the presence of identical ejective pairs obscures the restriction on

[10] This means that the phonotactic learning here happens over a sublexicon of roots; see §6.5 for more discussion.

[11] The use of a placeholder segment 'X' is of course not the ideal solution to this problem, as it obscures any other phonological generalizations that may hold of segments that are in an identity relation, like local restrictions on clusters of consonant-vowel interactions. A superior model would expand the search space of constraint to include algebraic notation. While Berent et al. (2012) present one potential method for constructing constraints of this type, no implementation of the model in that paper is available, nor has it been shown to be a general solution to phonological distinctions between identical and non-identical segments.

heterorganic pairs; a model without the 'X' transcription of identical ejectives learns no constraints on heterorganic ejective pairs, regardless of what projections are included. The simulations in this section show that when given the representational capacity to distinguish identical from non-identical ejective pairs (an essential property of any successful learner), this restriction is noticeable as a placeholder trigram in the baseline grammar, and motivates a nonlocal projection on which the phonotactic restriction can be stated.

4.3 Descriptive statistics and the baseline grammar

We first checked whether the baseline model finds the target placeholder trigram constraints that capture part of the nonlocal laryngeal phonotactics of the language: *[+plain][ ][+cg], *[+plain][ ][+sg, -continuant], and *[+cg][ ][+cg]. These constraints penalize illegal forms where the co-occurring consonants are separated by a single vowel like *[p'ak'a], but they do not extend to unattested consonant pairs separated by more material, e.g., *[p'ask'a]. The constraints are indeed found in the baseline grammar, when gain is set to 25 or below (a grammar of about 80 constraints). The model finds these constraints at a lower gain than in Quechua both because of the smaller size of the training set, and because each target constraint scopes over a smaller number of segments than the Quechua constraints. The fit of the baseline grammar to the test data is shown in Figure 4. This violin plot divides the harmony scores of CVCCV nonce words (left) and CVCV, VCV, VCCV nonce words (right, "other"). As in Quechua, the model makes the right distinctions between legal and illegal CVCV forms, since illegal CVCV forms violate the trigram constraints. It also correctly assigns higher scores to VCCV and VCV forms, some of which have just one laryngeal in onset position (e.g., [awk'a]. But the baseline model fails to distinguish legal nonce words from CVCCV words that violate laryngeal co-occurrence restrictions, assigning most of those forms relatively high scores.

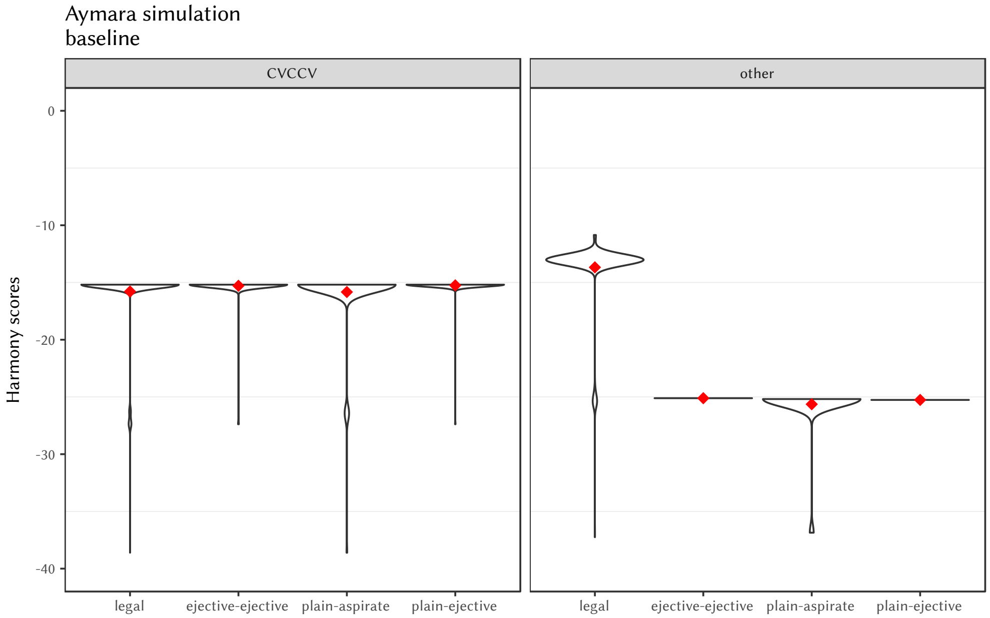

Figure 4: Aymara: harmony scores for nonce words, baseline grammar

Just as in Quechua, the placeholder trigram constraints are found in Aymara because laryngeal stops are frequent in both initial and non-initial positions, shown in Table 7. Likewise, onsets in Aymara are usually separated by just a single vowel, shown in Table 8. These properties of the language motivate placeholder trigram constraints against the underattested trigrams.

	plain	ejective	aspirate
initial	21%	21%	17%
non-initial	24%	7%	6%

Table 7: Aymara: frequency of plain, ejective and aspirate stops in initial and non-initial position. Percentages are out of all consonants, e.g., 21% of initial consonants are plain stops

Onset. . .onset ngrams	N of sequences	Proportion
CVCV	1316	66%
CVCCV	671	34%

Table 8: Aymara: Onset-V-onset trigrams as percentage of all onset...onset pairs (sequences in 1984 roots were counted)

4.4 Inducing projections and learning a final grammar

The three placeholder trigram constraints in the baseline grammar motivate two nonlocal projections, based on the natural class structure of the language. For the constraints *[+plain][ ][+cg] and *[+plain][

][+sg, -continuant], the smallest natural class projection is [-continuant, -sonorant], the oral stop projection. For *[+cg][ ][+cg], the smallest projection is [+cg]. When given these two projections, the model learns a final grammar that includes constraints against all the unattested sequences. The [-continuant, -sonorant] projection includes *[+plain][+cg] and *[+plain][+sg] and the [+cg] projection includes *[-wb][-wb], a constraint on any two segments on the projection of ejectives. The distribution of scores assigned to the test words is plotted in Figure 5. In the final model with two nonlocal projections, legal and illegal combinations of laryngeal stops are distinguished in all word shapes, (C)VCV and (C)VCCV.

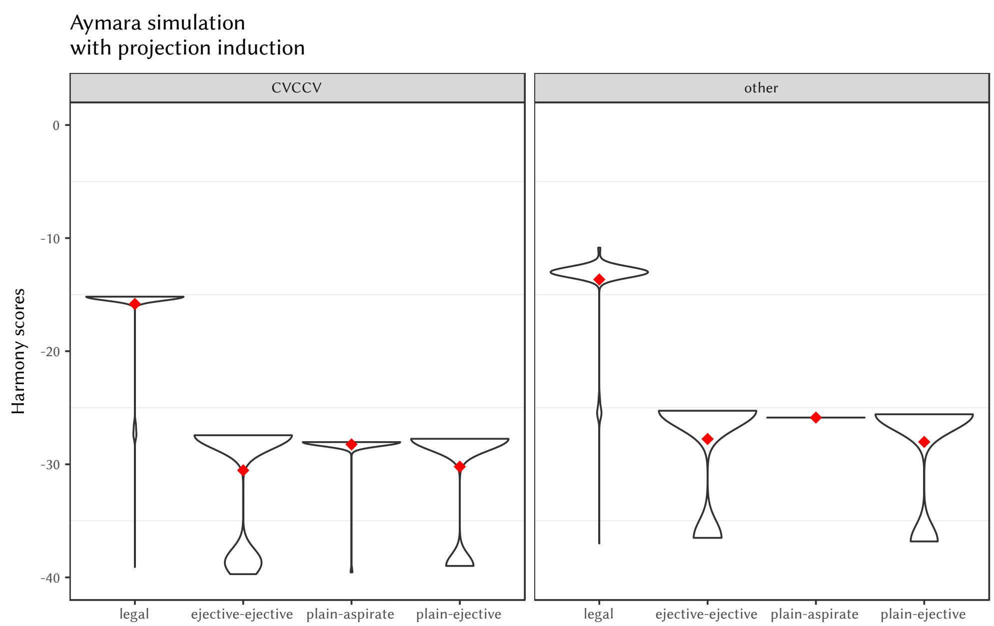

Figure 5: Aymara: harmony scores for nonce words, grammar with induced projections

Even though the grammar has a relatively low gain, the final grammar does assign worse scores to illegal forms than to the vast majority of legal forms. As can be seen in Figure 5, similar to Quechua, there are a small number of forms that were tagged as legal but are penalized by the grammar. This is due to the grammar including some trigram constraints against structures that are unattested in the data, but may or may not be accidental gaps. For example, the Aymara grammar includes a trigram constraint against #dental-mid vowel sequences, and a constraint against dental-mid vowel-labial sequences, both found in [nemq'e], one of the lowest rated legal nonce forms.

4.5 Summary

The Aymara case study builds on that of Quechua in two ways. First, it demonstrates that even constraints on smaller natural classes can be induced by attending to unattested trigrams in a baseline grammar; this is not specific to the broader restrictions in Quechua. Second, it presents a case where more than one nonlocal projection is motivated and kept in the grammar. In Aymara, the three restricted combinations could all be captured on one projection, [-continuant, -sonorant]. In our model, however, each placeholder trigram triggers a search through the smallest natural class projection motivated by that constraint. In Aymara, this means that both the [+cg] and [-continuant, -sonorant] projections are included in the final grammar and that the restriction on co-occurring ejectives is accounted for on a different projection than the restrictions on plain-ejective and plain-aspirate combinations. This is a good result, because a language may have multiple nonlocal restrictions that require different projections, even if the projections partially or fully overlap in what segments they contain.

5 Shona

Having shown how our projection induction procedure works from the baseline phonotactics of two languages with categorical laryngeal restrictions, we now turn to a somewhat different case: vowel co-occurrence restrictions in Shona. Shona shows both categorical and non-categorical restrictions on vowel height combinations, pieces of which are observable in the baseline phonotactics of the language. The baseline grammar for Shona reliably includes several placeholder trigrams referencing vowel height features, and these constraints motivate multiple nonlocal projections.

Shona provided the motivation for Hayes and Wilson's (2008) original argument that inductive phonotactic learning over *n*-grams requires nonlocal projections. Hayes and Wilson note that their baseline grammar for Shona finds placeholder trigram constraints that capture some of the restrictions on vowels, but in order to get the entire pattern, they give the [+syllabic] projection to the learner directly. In this section, we demonstrate that attending to these trigrams can be used to motivate projections, without the analyst supplying them to the learner. As we will show, the nature of the restrictions in Shona makes it hard for this particular learning model to arrive at a clean separation between harmonic and disharmonic forms. We discuss some remedies for this in §6.5 after showing what our learner induces without any modifications to the procedure.

5.1 Vowel height restrictions in Shona

Shona contrasts five vowels (see (8)), which are subject to phonotactic restrictions within verbal stems (Fortune 1980; Beckman 1997; Hayes and Wilson 2008; Mudzingwa 2010). According to Beckman (1997), the generalization is that [a] has an unrestricted distribution, but the mid vowels [e, o] can only occur in non-initial syllables if preceded by other mid vowels; furthermore, [e] cannot be followed by [i] but can be followed by [u], but [o] is generally followed by [o] rather than [u]. As we shall see, there are some exceptions to these generalizations.

(8) Vowel inventory of Shona, with features

	high	low	back
a	−	+	+
e	−	−	−
i	+	−	−
o	−	−	+
u	+	−	+

Since these restrictions are not categorical, we assessed the attestation of each vowel pair by computing the Observed/Expected ratio in a list of 4,600 verbal stems[12] compiled on the basis of the ALLEX corpus (Chimhundu 1996).[13] Table 9 shows Observed/Expected ratios for each ordered vowel pair. The Observed numbers track how often each vowel occurs as the first vs. second vowel in a two-syllable sequence, and the Expected numbers are calculated as the product of these positional probabilities divided by the total number of vowel pairs. If vowels are combining at random, the O/E ratio should be around 1; thus, the sequence [a. . .i] is slightly overattested. Looking along the diagonal (highlighted in gray), there is a clear preference for identical vowel sequences: each vowel is far more likely to be followed by the same vowel than by any other vowel, with O/E exceeding 1 for all identical vowel pairs. Importantly for a statistical learner, some combinations (boldfaced) are completely unattested or close to unattested: [a o], [a e], [e o],

[12]Morphologically, most of these stems appear to be imperatives, which are roots followed by some verbal projection suffixes (causatives, applicatives, etc.) and the [-a] suffix. Since all the citation forms of verbs end in [-a], this throws off the calculations for sequences that end in [a], so we removed that suffix for the purposes of O/E calculations. The suffix is present in the learning data for the simulations we report, however, since it is a categorical fact about Shona phonotactics that all words end in vowels.

[13]We opted to use a different corpus from Hayes and Wilson (2008), who used an incomplete scanned version Hannan (1974) that goes up to "m". Our corpus is slightly smaller but contains the full range of initial consonants, which matters for phonotactic learning. We verified that the distribution of vowel-vowel pairs is comparable in the two corpora.

[e i], [u o], [u e], [i e], [i o], [o i], [o u]. Other combinations are underattested, with O/E below 0.8: [e a], [e u], [i a], [i u], [o a], [u a].

	a	e	i	o	u
a	1.88	**0.062**	1.251	**0.0**	0.884
e	0.559	4.77	**0.009**	**0.0**	0.751
i	0.638	**0.019**	2.539	**0.030**	0.622
o	0.295	1.538	**0.092**	8.135	**0.025**
u	0.551	**0.006**	0.817	**0.0**	2.185

Table 9: O/E ratios for vowels in Shona verb stems

Given a vocalic projection, a phonotactic learner should be able to account for the restrictions with several bigram constraints: *o - highV, *high - midV, *e - o, *e - i, *a - midV. Note, however, that the statistical patterns are not as straightforward as simple height harmony: [e] and [o] do not pattern symmetrically, and neither do [i] and [u].

Though we are primarily interested in vowel co-occurrence restrictions as static phonotactics, restrictions on vowel height combinations are further supported by alternations. Shona verbal suffixes *-er/-ir, -es/-is, -ek/-ik,* and *-ew/-iw* alternate to match the height of preceding non-low vowels; the low vowel [a] conditions the appearance of [i] (see Table 10a). Fortune (1980, :21) discusses two suffixes with [u/o], which follow slightly different patterns. One of the round vowel suffixes is shown in Table 10b: its first vowel copies the stem vowel completely, and its second vowel alternates between [u/o]. Unlike these suffixes, verbal prefixes neither undergo nor trigger harmony (see Table 10c), and the final vowel suffixes [-e] and [-o] are also outside of the harmony system.[14] The failure of prefixes and final vowels to harmonize is not due to being external the phonological word, since unlike clitics, they count toward the disyllabic word minimum (Myers 1987; Downing and Kadenge 2015).

[14]Suffixes harmonize with verbal roots, but Fortune mentions a minor pattern whereby root vowels alternate to match the final -a or -e: [ndi-ger-e] 'I am seated' vs. [ku-gar-a], [ndi-ɲerer-e] 'I am silent' vs. [ku-ɲarar-a]. He lists five roots that follow this pattern; all alternate between [a] and [e] (Fortune 1980:20). We leave the phonological analysis of this for future work; for our present purposes, the important observation is that even the minor alternations are consistent with the phonotactic characterization of vowel harmony that affixes display.

a. Verbs: harmony in causative *-is/-es*, applicative *-ir/-er*, and extensive *-ik/-ek*			
-per-er-a	'end in'	-ip-ir-a	'be evil for'
-pofomadz-ir-a	'blind for'	-svetuk-ir-a	'jump in'
-om-es-a	'be dry'	-b͡vum-is-a	'make agree'
-taris-ik-a	'easy to look at'	-vereŋg-ek-a	'be numerable'
b. Verbs: harmony in the "un" suffix with rounded vowels			
-p͡fek-enur-	'undress'	-roj-onor-	'unwitch'
ʧat-anur-	'divorce'	-suŋg-unur-	'untie'
-piŋg-inur-	'unlatch'		
c. Verbs: prefixes and final vowels do not participate in harmony			
rim-is-ir-a	'make plow for!'	teŋg-es-er-a	'make sell for!'
mu-rim-is-ir-e	'make him/her plough for!'	mu-teŋg-es-er-e	'make him/her sell for!'

Table 10: Vowel patterns in Shona verbs (Fortune 1980, Downing and Kadenge 2015)

5.2 Methods: the training and testing data

The training data for our Shona simulations was the list of 4,600 verbal stems described above. To test the induced grammars, we generated a list of 10,000 pseudowords. The pseudowords were trisyllabic, started with a CV syllable, and ended in [a], like the verbs in our learning data. The middle syllable started with a singleton C (e.g., [m̤opera], orthog. *mh o p e r a*) or a CC that was robustly attested in the verb learning data (e.g., [d̤endowa], orthog. *dh e n d o w a*).[15] Each of the possible sequences of the five vowels [a e i o u] appeared in the first two syllables around 420 times. We classified the pseudowords into two categories: "disharmonic" and "harmonic". Disharmonic forms contain pairs of vowels that have near-zero attestation in the verb corpus and are described as disharmonic in phonological analyses of Shona (e.g., Beckman 1997).

5.3 The baseline grammar

The baseline grammar consistently includes several placeholder trigram constraints that penalize combinations of vowels across a single intervening segment. As for Quechua and Aymara, we focus on a model with relatively high gain and a small constraint size for illustrative purposes, as this represents a conservative hypothesis about what human learners have learned about their language. Table 11 lists the relevant placeholder trigrams found in a model with 170 gain (under 40 constraints) along with the vowel combinations these constraints penalize, and the smallest natural class based projection motivated by the

[15] The list of clusters we included: [gw, mw, bw, hw, kw, sw, nd, ŋg, mb, nz, nʤ].

constraint. These constraints penalize all of the disharmonic sequences, and also penalize one harmonic sequence, [e]-[u]. This combination is penalized because it contains a generally underattested natural class combination of a mid vowel followed by a high vowel.

	Constraint	Wt	sequences penalized		projection
			disharm.	harm.	
1.	[-high, -back][ ][-high, -low, +back]	13.709	eCo		[-high, -low]
2.	[+high][ ][-high, -low]	5.462	iCe, iCo, uCe, uCo		[-low]
3.	[+low][ ][-high, -low]	4.218	aCo, aCe		[-high]
4.	[-high, -low][ ][+high]	1.838	eCi, oCu, oCi	eCu	[-low]

Table 11: Shona verbs: constraints induced in the baseline run, the sequences they penalize and the projections they motivate.

While all the disharmonic sequences are penalized, the weight of constraints penalizing them varies greatly. The categorical constraint on eCo sequences has a high weight, since it is unviolated in the language. The other constraints all have relatively lower weights, because these constraints are not categorical and scope over sequences with varying degrees of attestation. Figure 6 shows the distribution of scores assigned by the baseline grammar to our testing set. Disharmonic vowel combinations that are separated by a single consonant are given a somewhat worse score than harmonic sequences, but no distinctions are made among vowel combinations that are separated by more than one consonant, since such structures do not fall under the scope of a trigram constraint.

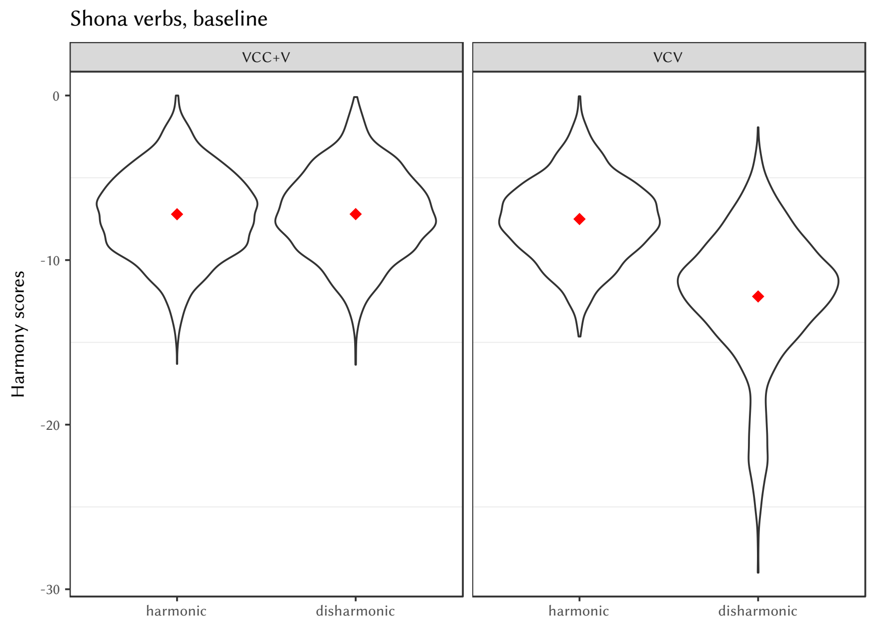

Figure 6: Shona: harmony scores assigned to nonce word test data by the baseline grammar, grouped by consonant strings separating the vowels.

Even in forms that contain just a singleton consonant, the scores assigned to disharmonic and harmonic forms overlap, reflecting the relatively low weights of constraints on vowel combinations (cf. the high constraint weights and good separation between legal and illegal forms in the Quechua and Aymara simulations). While the model for Shona includes the target placeholder trigrams, it also includes many other constraints on various, gradient phonotactic restrictions. The model here reflects the statistical support for the gradient restrictions on vowels, which the figure shows are not strong enough to create good separation between categories. This weak effect could be an accurate reflection of Shona speakers' judgments, or it could be that Shona speakers show somewhat stronger effects than are warranted by the statistical computation carried out by the model. The true strength of the vowel restrictions vs. other phonotactic generalizations is an empirical question that would require behavioral testing with Shona speakers. Regardless of the weight of constraints, the baseline model shows that dependencies between non-adjacent vowels are observable as placeholder trigrams, our main empirical point in this paper. The syllable structure of Shona allows these trigrams to be found because vowels are separated by just a single consonant a substantial proportion of the time, though longer strings of consonants are also possible between vowels. In Table 12, we show the numbers for how often vowels are separated by no consonants (VV), one unambiguous singleton consonant (VCV), two consonants (VCCV), or three consonants (VC-

CCV). We should note that the treatment of consonants in Shona is controversial; there are consonants with secondary articulation, and some phonologists analyze sequences as prenasalized stops, labialized stops, and so on (see, e.g., Mudzingwa 2010). If Shona is analyzed as having no consonant clusters, then 100% of vowels appear either in V. . .V bigrams or in VCV trigrams, and then nonlocal projections would not even be necessary for analyzing vowel co-occurrence. We assume that at least some of the consonant sequences are indeed clusters (see Maddieson 1990; Hayes and Wilson 2008; Stanton 2017a, ch. 2.4.3 for related discussion).[16]

V. . .V n-grams	Count	Proportion
V̲V̲	396	5%
V̲CV̲	6,333	79%
V̲CCV̲	1,232	15%
V̲CCCV̲	12	0.15%

Table 12: Shona: Vowel-to-vowel *n*-gram counts for the corpus of 4,688 verb stems.

5.4 Inducing projections and learning a final grammar

The four placeholder trigram constraints in the baseline grammar motivate three nonlocal projections that each pick out subsets of the vowels[17] in the language: [-high] ([a e o]), [-low] ([e i o u]) and [-high, -low] ([e o]). The final grammar includes all three projections and learns constraints on each one, summarized in Table 13. All disharmonic vowel combinations are penalized by some constraint, though the weight of the violated constraint varies, and some of these constraints also penalize some harmonic vowel pairs. Constraint 4 and the status of vocalic trigrams are discussed further in §5.5.

[16]The exhaustive list of clusters that occur in the Shona corpus: [ʦw, kw, ɗw, rw, mv, ʃw, zw, nw, tw, ŋw, jw, mw, ʣw, sw, hw, pw, ʒw, ɓw, gw, ɬw, nɗw, mb, nɗ, ŋg, nz, nʦ, ɲŋ, nj, ɗr, mbw, nhw, jŋ, ŋgw, nzw, nzv]. Many of these could be analyzed as labialized singletons or prenasalized stops or fricatives. The attractiveness of this move is somewhat tempered by the computational cost of increasing the number of natural classes. We do not know of a phonological analysis that would allow to treat sequences such as [nj] or [ɗr] as singletons.

[17]Technically, [-low] includes [e i o u j w], since we specified the glides in the feature set as [-syll] segments with vocalic features. When the feature set was rigged to exclude glides from vowel natural classes, the results did not change.

	constraint	projection	weight	sequences penalized disharm.	harm.
1.	*[-back][+back]	-high,-low	12.804	e-o	
2.	*[+low][-low]	-high	3.279	a-e, a-o	
3.	*[+high][-high]	-low	1.681	i-o, u-o, i-e, u-e	
4.	*[-high][+high,-back]	-low	1.674	e-i, o-i	e-a-i, o-a-i
5.	*#[-back][+back]	-low	1.408	#i-o, #e-o, #e-o	#e-u
6.	*#[+back]#	-high,-low	1.277	a-o, i-o, u-o, o-u	o-e, o-a
7.	*#[+syll,+back][-back]	-low	1.043	#o-i, #u-e	#o-e, #u-i, #o-a-i

Table 13: Shona verbs: constraints induced on induced projections, and the sequences they penalize.

The fit of the final grammar to the testing words is shown in Figure 6. Unlike the baseline grammar shown above in Figure 7, the grammar with nonlocal projections distinguishes vowel combinations across both a singleton consonant and longer consonant clusters.

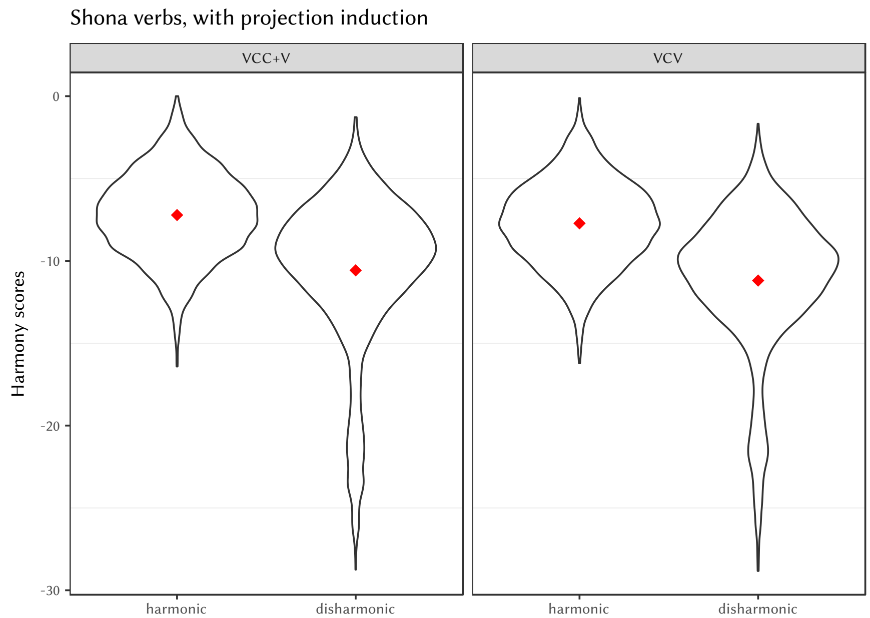

Figure 7: Shona: harmony scores assigned to nonce words by the final grammar after constraint induction.

As seen for the baseline grammar, the distinction between categories is smaller for Shona than for Quechua or Aymara, where there is little overlap between grammatical and ungrammatical testing words. This is an expected result, since the restrictions in Quechua and Aymara are categorical, while the restrictions in Shona are not. The weights of constraints on vowel combinations in Shona reflect the statistical support for each constraint in the training data. The constraint penalizing [e. . .o] combinations has a very high weight (12.8), consistent with there being zero violations of this constraint in the training data. For all of the other constraints, however, the restricted combinations of natural classes scope over combinations with some degree of attestation in the training data, and thus the weight of the constraint is lower.

Because of the amount of overlap in scores assigned to harmonic and disharmonic forms in the Shona grammars, we supplement the visualizations with statistical comparisons. We fitted two linear models with a dependent variable of MaxEnt harmony scores: one for the baseline grammar (see Table 14), and another for the grammar with projections produced by our learner (see Table 15). In both models, the scores of nonce words suffer when they have disharmonic vowel combinations, as seen in the negative coefficient estimates. But in the baseline model, nonce words with VCCV sequences actually receive higher scores compared to the linear model reference level (words with VCV scores and harmonic V combinations). In the final grammar, however, disharmonic V combos get a penalty regardless of whether they are in a VCV or a VCCV sequence (no significant interaction between disharmonic and VCCV).[18] Thus, the visuals reflect a statistically detectable difference between the two phonotactic grammars—the one with projections captures vowel harmony, not just the cooccurrence gap in VCV trigrams.

	Estimate	SE	t	Pr(>|t|)
Intercept	-7.51	0.05	-140.36	<0.000
disharmonic V combo	-4.70	0.08	-56.32	<0.000
VCCV	0.29	0.08	3.89	0.0001
disharmonic:VCCV	4.71	0.12	39.76	<0.000

Table 14: Linear model for MaxEnt harmony scores in Shona, baseline simulation

	Estimate	SE	t	Pr(>|t|)
Intercept	-7.72	0.06	-122.10	<0.000
disharmonic V combo	-3.47	0.1	-35.16	<0.000
VCCV	0.50	0.09	5.67	<0.000
disharmonic:VCCV	0.12	0.14	0.83	0.4

Table 15: Linear model for MaxEnt harmony scores in Shona, final simulation with induced projections

5.5 Multiple projections vs. one [+syllabic] projection

Finally, our induced projection grammar can be compared to one that a linguist would choose to analyze Shona—a grammar with a vowel projection. We ran a custom simulation with the [+syllabic] projection

[18] In both the baseline and the final grammar, VCCV forms receive slightly higher harmony scores than VCV forms. Since the constraints on CC sequences are poorly understood, we severely limited the range of clusters in our nonce words. This means that VCV forms, with their wider range of consonants in medial position, are more likely to violate bigram constraints on CV and VC sequences. We do not know what the status of these constraints is in Shona speakers' grammars, so it is an open question whether the computational learner is overfitting.

and tested it on the same nonce words, and the results are shown in Fig. 8. The plot shows the same trend for disharmonic words to have lower scores than harmonic ones, but this simulation does not manage to make a categorical separation between them. Just like our mosaic projection grammar, this one finds all the constraints against disharmonic forms, but it also finds some constraints against harmonic ones (e.g., *#[-high,-low][+low], which penalizes [e-a] and [o-a], and *#[+high,-back][+high,+back], which penalizes [i-u]).

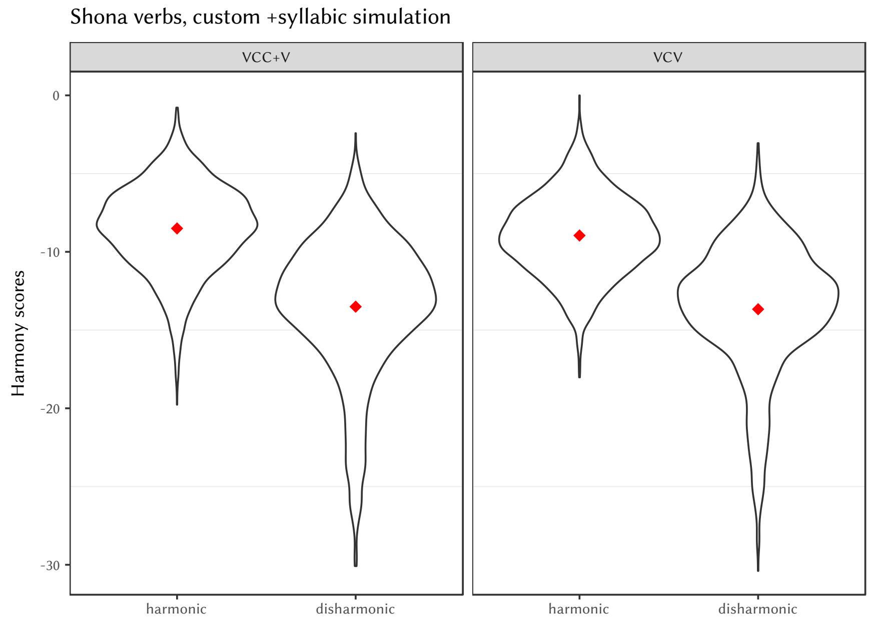

Figure 8: Shona with a manually supplied [+syll] projection.

One area where a grammar with a [+syll] projection would be expected to be more accurate than a projection grammar is in dealing with opacity. An opaque segment Z prevents normally restricted segments X and Y on either side of it from agreeing with respect to a feature; thus, X. . .Y is not allowed, but X. . .Z. . .Y is. If the segment Z is not present on the projection that includes X and Y only, then constraints on the X-Y projection will incorrectly rule out X-Z-Y. Our inductive learner in fact met with this problem when it induced Constraint 4 on the [-low] projection, which excludes the vowel [a]. This vowel is described as opaque with respect to height harmony (Beckman 1997; Hayes and Wilson 2008), but as it turns out, the statistical support for its opacity is rather weak (see Table 16). The disharmonic sequences [e. . .i], [o. . .i] and [o. . .u] are indeed underattested compared to height-agreeing sequences (recall Table 9), but so are trigrams of these vowels separated by [a]—there are only 51 examples of such vowel trigrams. O/E values for trigrams are generally lower than for bigrams, since the joint probability of finding a sequence of three vowels is lower than that of two. The formula for positional O/E trigram frequency is: $\frac{N(abc)}{N(a)*N(b)*N(c)/N_{trigrams}}$, where *N(a)* is the frequency of *a* as the first segment in a trigram, *N(b)* is the frequency of *b* as the second segment in a trigram, and so on. $N_{trigrams}$ is the total number of trigrams on the relevant projection—in the list of Shona verbs, there are 5360 trigrams. As shown in Table 16, the raw frequencies of trigrams that supply the evidence for the opacity of [a] are low compared to frequencies of identical vowel trigrams or mixed-height trigrams. Most importantly, the O/E for opaque trigrams is hard to distinguish from that of some disharmonic trigrams—-for example, [a e a] occurs more often than [o a i], which is supposed to be good, and has a higher O/E than [o a u] (non-zero, although minuscule). If all the learner has to go on is the comparatively high O/E of the sequence [e a i], then the evidence for the goodness of mid-low-high opaque sequences is not strong.[19]

identical	N	O/E	mixed	N	O/E	opaque	N	O/E
a a a	290	0.0002	u u a	505	0.0004	o a u	0	0.0000
e e e	155	0.0055	o o a	319	0.0015	o a i	11	0.0001
o o o	96	0.0178	e e a	347	0.0008	e a i	40	0.0003
u u u	197	0.0011	u i a	216	0.0002			
i i i	121	0.0008	a e a	16	<0.0000			

	as V1	as V2	as V3
a	1874	1108	3830
e	883	733	232
i	788	1420	687
o	656	459	96
u	1159	1640	515

Table 16: Positional O/E calculations for trigrams on the vocalic projection in Shona verbs: weak evidence for opacity of [a] to harmony. The subtable on the right shows the positional frequency of each vowel in vocalic trigrams.

We conclude that Shona does not supply clear-cut quantitative evidence for a vocalic projection: it is possible to approximate the generalizations about vowels on projections that include only subsets of vowels, as shown by comparing our mosaic projection grammar to the grammar with a manual [+syllabic] projection. The nature of the learning data makes it difficult for a statistical learner to match the generalizations that linguists formulate about this language, regardless of the projections that it has access to. Opacity may be better noticed in Shona by looking at morphological alternations, a broader point we return to in 6.5.

[19] An anonymous reviewer suggests evaluating the fit of the [+syllabic] phonotactic grammar with that of our mosaic grammar in a linear model, as we did for the baseline vs. mosaic grammars earlier. Unsurprisingly, given the visual impression in the plot, there is a significant effect of vowel harmony status on harmony scores in a linear model for the [+syllabic] grammar. The question, then, is whether it is possible to decide which model is better on the basis of such statistical comparisons. The usual methods of model comparison such as Akaike Information Criterion do distinguish these models, favoring [+syllabic] over the mosaic model (52,773 vs. 53,140—lower is better)—but this comparison also favors the baseline model (AIC=49,724) over both of the models that capture the vowel harmony generalizations that we are after. The statistical method of evaluating models therefore points away from linguistic intuitions, which could be a potential problem for us. The only way to find out which model captures the right generalizations is to test them experimentally on human speakers of Shona.

5.6 Summary

As in Quechua and Aymara, a baseline grammar looking only at the linear string of segments finds placeholder trigram constraints that penalize all of the restricted vowel combinations in Shona. The Shona case is different from Quechua and Aymara in several ways, underscoring the generality of our proposal. Many of the restrictions in Shona are noncategorical, and accounting for the distribution of Shona vowels requires multiple constraints on smaller classes of segments than in either Quechua or Aymara. In Shona, the trigram placeholder constraints motivate three distinct projections on subsets of interacting vowels. A phonologist would be more likely to postulate a single projection that includes all the vowels in Shona, but it's not clear how this projection could be learned from a baseline grammar without projections. Our model incorporates a simple hypothesis about nonlocla projections: only the classes that are referenced by a baseline placeholder trigram constraint are projected. While an analysis with a single projection may be formally more elegant, our model with multiple projections still captures the distribution of vowels and distinguishes harmonic and disharmonic forms in much the same way as a single projection does.

6 General discussion

We've shown through three case studies that nonlocal phonological interactions that hold at arbitrary distances may be observable as underattested trigrams in the linear string. For languages with this property, we've proposed a simple method of using a placeholder trigram constraint in the baseline grammar to construct a nonlocal projection that allows the grammar to fully capture the nonlocal interaction. In this section, we first relate our proposal to previous work and then go on to discuss the place of our model in accounting for nonlocal restrictions more generally. We address potential cases where our model may over- and under- generalize from placeholder trigrams, as well as discussing opacity and blocking patterns, the interaction between our model and syllable structure, and the role of morphological alternations.

6.1 Previous computational and theoretical work

Ours is not the first attempt to induce nonlocal phonological constraints from learning data. In contrast to our approach, Heinz (2010), Jardine (2015) and Jardine and Heinz (2016) characterize nonlocal phonology as an idealized problem of searching for unattested substrings. Their learners memorize attested precedence relations between segments and induce constraints against those sequences that they have not encountered. One of the problems with this approach is that it can reify accidental gaps to the level of categorical phonotactic constraints, whereas stochastic patterns with exceptions will stymie it (Wilson and Gallagher 2018).[20]

These models have also been motivated and evaluated only in the form of theoretical proofs over idealized data sets, and have not been tested on natural language data. We suspect that were they to be implemented, they would run into some of the problems we discussed in §2.4, since at least some of the algorithms involve searching for nonlocal n-grams. The approaches in particular will run into problems when seen in the context of the larger problem of learning phonotactics. A non-idealized learner is solving multiple problems at the same time: non-local dependencies alongside local phonotactics, and word and morpheme segmentation (Adriaans and Kager 2010). There is acquisition research and learnability arguments that segmentation interacts with nonlocal dependencies (Kastner and Adriaans 2017; Van Kampen et al. 2008 and the references therein). If word boundaries are not known, then a nonlocal substring learner's window expands far beyond the 25-segment size maximum of our corpus-based plots in Fig. 1. We do not see an easy way to combine nonlocal substring computations with segmentation, since even nonlocal bigram calculations will get out of hand when the strings get very long, and when the bigrams are taken to consist of natural classes (as opposed to segments). On the other hand, an approach such as ours can be combined with learning segmentation, since it hinges on the properties of local n-grams only.

Futrell et al. (2015) propose a very different approach—their learner is statistical and uses features, keeping track of local and non-local n-grams. The approach to nonlocal phonology searches for co-occurrence constraints by traversing a feature geometry tree. As long as the search through the tree proceeds directionally, it becomes a subcase of the Directed Acyclic Graph problem, which has well-known algorithmic solutions. When the learner is tested on a variety of transcribed dictionary corpora, it finds vowel harmony tendencies in languages like Turkish, but it also identifies harmony patterns in languages that do not have any (this is not a damning critique since statistical learners are generally guilty of finding patterns that linguists consider accidental). Their model is tested on held out forms, which are all phonotactically legal, not a large set of legal and illegal nonce words, so it is not clear how well the resulting grammar distinguishes ungrammatical novel forms from licit, held out forms. We are also skeptical of the assumption that all

[20] A similar criticism can be applied to the model of Goldsmith and Riggle (2012). They argue that their model discovers the projection relevant to Finnish vowel harmony, but it does so over segmental rather than featural representations—thus, the comparison is between V-to-V vs. V-to-C nonlocal relations. This assumes that the learner is considering only V and C natural classes, thereby giving the learner a vocalic projection for free. It also allows the learner to notice accidental nonlocal co-occurrence restrictions that do not involve segments from the same natural class, which our learner cannot detect.

nonlocal restrictions can be characterized using a feature geometry; in particular, most structured geometries cannot gracefully capture patterns that involve features from different branches of the feature tree. A more flexible approach would allow the learner to identify the relevant natural classes from language evidence—and we demonstrate that our learner has this capability.

6.2 Over- and under-generalization of the placeholder trigram approach

We have demonstrated that three different languages with nonlocal restrictions exhibit these restrictions as placeholder trigram constraints in the baseline grammar. To further support our approach, we should understand the conditions under which a nonlocal pattern may not be observable as a placeholder trigram on the baseline grammar—that is, the conditions under which our proposal would undergeneralize. We should also consider whether our model is likely to uncover nonlocal interactions in languages that are not typically thought to have them—that is, whether our model overgeneralizes and reifies accidental gaps. A full exploration of both over- and under-generalization would require a broad empirical survey and detailed mathematical investigation of the properties of the Hayes and Wilson model, which we leave for future work. Here, we summarize our observations from working with the three languages in our case studies, as well as a few others.

We have in fact identified several cases of undergeneralization, where a nonlocal restriction is not reflected in the baseline grammar as a placeholder trigram constraint. The Hayes and Wilson model has several properties that make finding some placeholder trigrams a challenge. First, the model prefers shorter constraints, so all grammars, regardless of gain or constraint set size, contain relatively few trigram constraints compared to bigram constraints. Second, the model prefers constraints that use fewer features, so constraints on natural classes that require two or three features to define may be missed. A third factor, which interacts with the first two, is that the model does not learn exhaustively. Instead, the baseline grammar assesses whether adding a constraint to the grammar significantly improves the fit to the training data. Even some categorical trigram constraints may not have enough statistical support (that is, the absence of such trigrams may not be sufficiently statistically surprising) to be included in the grammar.

We observed this problem in Quechua, when looking at just the set of roots (as opposed to phonological words, as reported in §3). While the baseline grammar always includes placeholder trigram constraints that hint at the laryngeal restrictions in each language, Quechua also has a categorical restriction on uvulars and velars coocurring in roots, and this restriction is not consistently reflected in placeholder trigrams.

One possible reason that these constraints are absent is because many unattested combinations of uvulars and velars also fall under the purview of the laryngeal restrictions. A hypothetical constraint *[dorsal, +RTR][][dorsal, -RTR] would penalize just three combinations that don't violate the laryngeal restrictions—[qVk], [q^hVk], [q'Vk]—and six forms that also violate a broader restriction on laryngeal combinations—[qVk'], [qVk^h], [q^hVk'], [q^hVk^h], [q'Vk'], [q'Vk^h]. The work of constraints on uvular-velar cooccurrence largely duplicates the work of laryngeal constraints in the grammar, and since laryngeal constraints hold of larger classes of sounds, these constraints are preferred. Whether or not the model finds constraints on uvulars and velars also depends on feature specification. If uvulars and velars are specified as dorsal, and distinguished by [RTR], then each class requires two features to pick out and the model is less likely to include constraints on these classes. If instead two privative features are used, [velar] and [uvular], the model is more likely to use these single-feature classes in constraints.

The importance of picking out a natural class with a single feature was echoed for the laryngeal restrictions in Aymara (§4) and some preliminary work on stridents in Kinyarwanda. For Aymara, the restrictions on plain-ejective and plain-aspirate combinations are only found if plain stops can be picked out using a single feature. Our model finds these constraints because it is given the privative feature [plain]; if, instead, the model is given binary [cg] and [sg] and plain stops must be picked out with two features as [-cg, -sg], then the model fails to learn anything about the distribution of plain stops. In Kinyarwanda, retroflex and dental sibilants interact, but these restrictions are only found by the grammar if the feature set defines these classes with a single feature, e.g., with the ad hoc features [retroflex strident] and [dental strident]. If these classes are instead described using three binary features as [+coronal, +strident, -distributed] and [+coronal, +strident, +distributed], the model does not include constraints that reference these classes. This finding reflects a quirk of the Hayes and Wilson model that should likely be reconsidered. Many phonological patterns involve natural classes that may require many features to be picked out, and indeed the same pattern may require a different number of features in different languages. We further anticipate that the model may have difficulty finding placeholder trigram constraints that refer to very small natural classes, e.g., interactions between just two segments like [l] and [r], regardless of how many features are required to pick out these classes.

While our explorations with the Hayes and Wilson learner have shown that undergeneralization is a serious concern under certain conditions, we have not found any problematic instances of overgeneralization of baseline trigram constraints. Since the model is fairly conservative about positing trigram constraints, even models with low gain and a large constraint set contain relatively few placeholder trigram constraints. Should a baseline grammar include a placeholder trigram constraint that does not correspond to a robust nonlocal interaction in the language, our model will deal with this the same way it deals with other accidental gaps: by searching this projection for constraints that meet the gain criterion of the model. If a placeholder trigram corresponds to an accidental gap, as opposed to being a local instantiation of a broader restriction, the search through a projection will not find many useful constraints, or will only find constraints that it assigns a low weight. When our model is run on Russian, for example, it finds no placeholder trigrams—no matter the gain or constraint number.[21] Since Russian is not known to have any strong nonlocal phonological interactions, this is to be expected, and we consider it a good outcome.

Perhaps more interesting is the failure of the model to find any placeholder trigrams in a corpus of Mongolian, a language with vowel harmony (Walker 2001; Svantesson et al. 2005). The An Crúbadán corpus of Mongolian (available at http://crubadan.org/) supplies evidence for vowel harmony—the learner finds multiple vowel harmony constraints when given the [+syllabic] projection directly. But it does not find anything like it on its own. The reasons for this require investigation in future work, but we can speculate. First, Mongolian does have a large number of segmental contrasts, and its syllable structure is too varied for our learner. Vowels are separated by two or more consonants almost half the time, so there are far fewer VCV trigrams than Shona. Second, Mongolian also has many local CV segmental interactions, and just accounting for those requires so many bigram constraints that Hayes and Wilson's heuristics deprioritize trigrams. But Mongolian also has extensive vowel alternations, so its vowel harmony could be captured in a different way (see §6.5).

In sum, we have identified cases where our placeholder trigram method fails, but no cases where it falls prey to accidental gaps that affect the overall fit of the model. In future work, we plan to explore more thoroughly the circumstances under which a nonlocal phenomenon will be observable as a local trigram, as well as other types of evidence that may be available to a learner in languages where nonlocal interactions are not learnable as a placeholder trigram (see §6.5 below for a discussion of the role of morphological alternations). Overgeneralization is handled in our model in the same way it is handled in Maximum En-

[21] Russian is one of the languages that causes the Java implementation of the learner to run out of memory at the constraint enumeration stage, due to the large number of natural classes. We got around this for Russian by redefining the feature set to use several privative oppositions and not transcribing certain important phonotactic patterns (such as vowel reduction). This reduces the number of natural classes for the learner to deal with, and with it the ability to make certain phonological generalizations. Even this move did not help with Hungarian.

tropy models in general—by only learning constraints (on the baseline projection or a nonlocal projection) that satisfy the gain criterion.

6.3 Opacity and blocking

A second type of challenge for our model are systems that show opacity or blocking. Our learner builds projections defined by the smallest natural class that includes both natural classes mentioned in the placeholder trigram constraints, a procedure that is simple, deterministic, and represents the maximally general hypothesis about the transparency of intervening segments. This smallest projection will be the correct one, unless the language has opaque or blocking segments. This problem was shown in Shona, where placeholder trigram constraints motivated projections on subsets of vowels, but not a [+syllabic] projection containing all vowels. Without such a projection, the opacity of a low vowel to interactions between high and mid vowels cannot be captured, because no projection includes both the low vowel and the interacting high and mid vowels. A similar problem is found in a cursory look at how our model handles sibilant harmony in Kinyarwanda. Walker et al. (2008) report that dental sibilants may not be followed by retroflex sibilants in the next syllable ([ⁿ̢ʂa:ʐe] 'I am old', *[ⁿsa:ʐe]). At further distances, harmony is generally optional ([-saka:ʐe] ∼ [-ʂaka:ʐe] 'cover the roof with, perf.'), but is impossible just in case a non-sibilant coronal or palatal intervenes ([sí:ta:ʐe] 'make stub, perf.' *[ʂí:ta:ʐe]. When our model is trained on a dictionary word list of 2576 forms (Cox et al. 1998), it finds a placeholder trigram that represents the harmony restriction, penalizing dental sibilant-X-retroflex sibilant sequences. This trigram constraint motivates a strident projection, which includes the dental and retroflex sibilants but not the opaque consonants. The model can thus learn a constraint on a nonlocal projection that enforces harmony at greater distances, but the resulting grammar will not be sensitive to the identity of interening consonants and thus will not capture opacity.

For both Shona and Kinyarwanda, there is some question as to how much statistical support there is for opacity. In Shona, we showed that even in our large set of forms there is little statistical support for the underattestation of the trigrams that could show opacity. In Kinyarwanda, the dictionary word list contains zero forms that show opacity, and a quick look at the web corpus on An Crúbadán finds fewer than 10 such forms in a list of 50,000. If a clear, statistically supported case of opacity can be found as a phonotactic system, our model will need to be modified to learn larger projections just in case they are necessary. We leave development of the model in this direction for future work, pending identification of a statistically robust case of opacity.

6.4 Predictions for syllable structure and distance effects

Our model's main distinguishing trait is that it is driven by language-specific characteristics that are observable from baseline phonotactics, without projections. A simple trigram-based learner identifies constraints that govern segmental co-occurrence across an irrelevant constituent—which is the definitional property of a nonlocal phonological interaction. Our learner detects the presence of such placeholder trigram constraints in the baseline grammar and isolates natural classes involved in the interaction, searching projections in a systematic way for constraints that are motivated in the language. This procedure is inspired by old insights from phonological research: that segments interact with each other nonlocally when they are part of a natural class (McCarthy 1986; Rose and Walker 2004 and others), and that nonlocal interactions are easier to notice in languages where consonant and vowel arrangements are templatic (McCarthy 1989) than in languages where syllable structure is more complicated and unpredictable. In our view, the connection between these properties receives a learning-theoretic explanation and opens up a line of future research.[22] By attuning only to interactions that are observable in a local trigram, and constructing the smallest natural class based projection from such a trigram, our model avoids the computational cost of an exhaustive search and also reduces the likelihood of finding accidental gaps.

Our proposal may also contribute to the explanation for a well-known feature of nonlocal restrictions: *distance effects* (Cohn 1992; McCarthy 1994; Suzuki 1998; Berkley 2000; Hansson 2001; Rose and Walker 2004; Frisch et al. 2004; Albright and Hayes 2006; Hayes et al. 2009; Kimper 2011; Berkson 2013; Bennett 2015; Stanton 2017b). In distance effects, the nonlocal restrictions hold more strongly across one intervening segment, and weakly or not at all when the segments are separated by more material. Our approach offers a different characterization of these effects: the restricted sequence is penalized by a baseline placeholder trigram, but the learner has either failed to find evidence for the relevant projection or finds the

[22] An anonymous reviewer asks why nonlocal interactions aren't more frequent in Polynesian languages, which have very simple syllable structure. First, several languages of the region have been noted for their nonlocal consonant interactions (see Blust (2012) for a review of OCP effects in these languages, as well as Coetzee and Pater 2008; Zuraw and Lu 2009). While we do predict that nonlocal interactions should be learnable via our method in Polynesian languages, there may be other reasons, including chance, why a language does or does not exhibit a particular type of pattern. For example, in a language with a small segmental inventory and simple syllable structure, nonlocal phonological dependencies introduce additional limitations on possible words, resulting in a relatively small set of unique words, unless words are extremely long. Morphological reduplication may make phonotactic nonlocal dependencies difficult to detect, since patterns may be ambiguous between a phonotactic and a morphological analysis.

evidence inconsistently. If this is on the right track, then we may have a learnability explanation for distance effects.

6.5 Learning nonlocal projections from alternations

One factor that we did not address but is likely crucial to learning some of the more complicated nonlocal interactions is that they are morphologically restricted: they are either evinced in affixal alternations or hold as static morpheme structure constraints over roots (see Rose and Walker 2004 for in-depth discussion). Indeed, the patterns must be one or the other to be observable as phonotactic constraints. In Quechua, laryngeal co-occurrence constraints hold over morphologically complex words without alternations. There are thus two types of morpheme structure constraints in the language: (i) no ejectives/aspirates in affixes, and (ii) the various co-occurrence constraints on the stop projection in roots. The simulation we reported in §3 used morphologically complex words as learning data, but the evidence for nonlocal restrictions is much more concentrated if the learner is given a list of roots instead. In Aymara—a language that is minimally different from Quechua—the constraints hold only of roots and are violated in words with affixes, which do have ejectives and aspirates. In order to learn the generalizations about Aymara roots, the learner presumably separates roots into their own group, a *sublexicon*, for phonotactic learning (Gouskova and Becker 2013; Becker and Gouskova 2016; Gouskova et al. 2015; Becker and Allen submitted).

Our simulation for Shona (as well as Hayes and Wilson's 2008 simulation) implicitly assumed that phonotactics are learned over sublexicons: our training data were citation forms of verbal stems, the only place where vowel co-occurrence restrictions hold. The nouns of Shona do not respect these phonotactics, and other morphological forms of verbs violate them as well (Fortune 1980). When we trained the learner on the entire ALLEX word corpus (Chimhundu et al. 1996), the baseline grammar did not include any placeholder trigram constraints, so the learner did not induce any projections. When given a vocalic projection directly, the learner found trivial constraints (e.g., *##, "words should have a vowel") and low-weighted constraints on rare trigrams (e.g., *[+high,+back][-high, -back][+high,+back], with a weight of 0.65). It did not make any distinctions among harmonic and disharmonic nonce words in the test set, either (Welch's Two Sample t-test, VCV harmonic vs. VCV disharmonic t(4300)=0.77, p=0.4, VCCV harmonic vs. VCCV disharmonic, t(4300)=-1.7, p=0.08).

We hypothesize that in cases where alternations enforce the restrictions, these alternations are also key to identifying the right projections. Alternations help in three ways.

First, alternations make the restriction highly salient. They present the learner with a clear problem to solve: what is responsible for the systematic mismatch between the different forms corresponding to the same meaning? Both linguists and human learners attend to alternations, so they offer a shortcut to the difficult problem of noticing the presence of nonlocal interactions when the language does not otherwise cue them in its local phonotactics.

Second, when learning phonotactics over sublexicons, the learner has access to concentrated evidence where certain sequences will be overattested and others will be underattested or unattested. This was the case in our Shona verb stem training set, and is a general characteristic of sublexicons (cf. the so-called "islands of reliability"—near-inviolable generalizations about morphophonologically defined classes, Albright 2002; Albright and Hayes 2003; Becker et al. 2011; Gouskova et al. 2015; Becker and Gouskova 2016).

Third, the disparities between the allomorphs can be a guide to the relevant projection. For example, in Shona, the applicative alternates between *-ir/-er*, and the "un-" morpheme alternates between *-onor/-unur/-enur/-anur/-inur*. If the alternation cannot be attributed to segmentally local conditioning, a projection could be formed by collecting the non-matching segments [o, u, e, a, i] and finding a natural class that includes all of them—here, [+syllabic]. In order to work for nonlocally conditioned alternations with opacity, the procedure would have to be more elaborate; we leave this for future investigation.

The entire learning trajectory could then start with segmentally local baseline learning over phonological words only, as for Quechua. Once the learner becomes morphologically aware, learning would proceed to an automatically created sublexicon for roots; this would be necessary for languages like Aymara. Finally, local and nonlocal alternations would be sorted out, and if local conditioning does not explain alternations, projections would be tested. The learner does not know a priori whether the alternations are nonlocally conditioned or even phonologically conditioned (the pattern could be lexically conditioned suppletion, after all), so this kind of learning should be harder and will happen at a later stage.

7 Conclusion

We presented an inductive learning model that capitalizes on the observation that nonlocal phonological interactions are segmentally almost local at least some of the time—that is, they can be observed by keeping track of segmental trigrams whose medial segment is phonologically a placeholder, *X[ ]Y. We demonstrated that the full extent of nonlocal interactions can be captured by positing a representational projection for the smallest natural class that includes X and Y, which incorporates the most general hypothesis that all but the interacting segments are irrelevant to the restriction. Our learner identified the correct generalizations about laryngeal co-occurrence constraints on consonants in Quechua and Aymara, and it also found the vowel co-occurrence restrictions in Shona (though the weight of the constraints did not allow for a good separation of harmonic and disharmonic nonce forms).

While we do not think that this is the final word on learning nonlocal phonological interactions, this kind of learning offers a plausible starting point in a framework that does not assume that the learner has access to universally available projections. Instead, the learner attends to the properties of the language, and is moved to posit projections only when encountering certain kinds of evidence. We see this proposal as a promising avenue for tackling the considerable search space of nonlocal interactions in a structured way.

References

Adriaans, Frans, and René Kager. 2010. Adding generalization to statistical learning: The induction of phonotactics from continuous speech. *Journal of Memory and Language* 62:311–331.

Albright, Adam. 2002. The identification of bases in morphological paradigms. Doctoral Dissertation, University of California, Los Angeles, Los Angeles, CA.

Albright, Adam, and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90:119–161.

Albright, Adam, and Bruce Hayes. 2006. Modeling productivity with the Gradual Learning Algorithm: The problem of accidentally exceptionless generalizations. In *Gradience in Grammar: Generative perspectives*, ed. Gisbert Fanselow, Caroline Fery, Matthias Schlesewsky, and Ralf Vogel, 185–204. Oxford: Oxford University Press.

Becker, Michael, and Blake Allen. submitted. Learning alternations from surface forms with sublexical phonology. *Phonology*, URL `http://ling.auf.net/lingbuzz/002503`.

Becker, Michael, and Maria Gouskova. 2016. Source-oriented generalizations as grammar inference in Russian vowel deletion. *Linguistic Inquiry* 47:391–425.

Becker, Michael, Nihan Ketrez, and Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish devoicing neutralization. *Language* 87:84–125.

Beckman, Jill. 1997. Positional faithfulness, positional neutralization, and Shona vowel harmony. *Phonology* 14:1–46.

Beckman, Jill. 1998. *Positional faithfulness.* New York: Routledge.

Bennett, William G. 2015. Assimilation, dissimilation, and surface correspondence in Sundanese. *Natural Language & Linguistic Theory* 33:371–415.

Berent, Iris, Colin Wilson, Gary Marcus, and Doug Bemis. 2012. On the role of variables in phonology: Remarks on Hayes and Wilson (2008). *Linguistic Inquiry* 43:97–119.

Berkley, Deborah. 2000. Gradient OCP effects. Doctoral dissertation, Northwestern University.

Berkson, Kelly Harper. 2013. Optionality and locality: Evidence from Navajo sibilant harmony. *Laboratory Phonology* 4:287–337.

Blust, Robert. 2012. One mark per word? Some patterns of dissimilation in Austronesian and Australian languages. *Phonology* 29:355–381.

Chimhundu, H. 1996. *Duramazwi reChiShona.* Harare: College Press Publishing Ltd. URL `http://www.edd.uio.no/`.

Chimhundu, Herbert, Oddrun Grønvik, Christian Emil Smith Ore, and Daniel Ridings. 1996. The African Languages Lexicon Project (ALLEX). URL `http://www.edd.uio.no/allex`.

Coetzee, Andries W., and Joe Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory* 26:289–337.

Cohn, Abigail. 1992. The consequences of dissimilation in Sundanese. *Phonology* 9:199–220.

Colavin, Rebecca S, Roger Levy, and Sharon Rose. 2010. Modeling OCP-Place in Amharic with the Maximum Entropy phonotactic learner. In *Proceedings of the 46th meeting of the Chicago Linguistics Society*, volume 46, 27–41. Chicago, IL: Chicago Linguistic Society.

Cox, Betty Ellen, Myra Adamson, and Muriel Teusink. 1998. Kinyarwanda-English Dictionary.

De Lucca, M. 1987. Diccionario Práctico Aymara-Español, Español-Aymara. *Editorial Los Amigos del Libro, La Paz* .

Della Pietra, S., V. Della Pietra, and J. Lafferty. 1997. Inducing features of random fields. *IEEE transactions on pattern analysis and machine intelligence* 19:380–393.

Downing, Laura J, and Maxwell Kadenge. 2015. Prosodic stems in Zezuru Shona. *Southern African Linguistics and Applied Language Studies* 33:291–305.

Fortune, George. 1980. *Shona Grammatical Constructions Vol. I*. Harare: Mercury Press, 2nd edition edition.

Frisch, Stefan A., Janet B. Pierrehumbert, and Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22:179–228.

Futrell, Richard, Adam Albright, Peter Graff, Edward Gibson, and Timothy J O'Donnell. 2015. A Probabilistic Autosegmental Model of Phonotactics. Ms., MIT.

Gafos, Adamantios. 1999. *The Articulatory Basis of Locality in Phonology*. New York: Garland.

Gallagher, Gillian. 2014. An identity bias in phonotactics: Evidence from Cochabamba Quechua. *Laboratory Phonology* 5:337–378.

Gallagher, Gillian. 2015. Natural classes in cooccurrence constraints. *Lingua* 166:80–98.

Gallagher, Gillian. 2016. Asymmetries in the representation of categorical phonotactics. *Language* 92:557–590.

Gallagher, Gillian, and Jessica Coon. 2008. Distinguishing total and partial identity: Evidence from Chol. *Natural Language and Linguistic Theory* 27:545–582.

Goldsmith, John, and Jason Riggle. 2012. Information theoretic approaches to phonology: The case of Finnish vowel harmony. *Natural Language and Linguistic Theory* 30:859–896.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. In *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, ed. Jennifer Spenader, Anders Eriksson, and Östen Dahl, 111–120. Stockholm: Stockholm University.

Gouskova, Maria, and Michael Becker. 2013. Nonce words show that Russian yer alternations are governed by the grammar. *Natural Language and Linguistic Theory* 31:735–765. Available on Lingbuzz at `http://ling.auf.net/lingBuzz/001456`.

Gouskova, Maria, Sofya Kasyanenko, and Luiza Newlin-Łukowicz. 2015. Selectional restrictions as phonotactics over sublexicons. *Lingua* 167:41–81. URL `http://ling.auf.net/lingbuzz/002673`.

Hannan, Michael. 1974. *Standard Shona Dictionary*. Harare: College Press in conjunction with the Literature Bureau.

Hansson, Gunnar Olafur. 2001. Theoretical and Typological Issues in Consonant Harmony. Doctoral Dissertation, University of California, Berkeley, Berkeley.

Hardman, Martha James. 2001. *Aymara*, volume 35. München: Lincom Europa.

Hayes, Bruce, and James White. 2013. Phonological naturalness and phonotactic learning. *Linguistic Inquiry* 44:45–75.

Hayes, Bruce, and Colin Wilson. 2008. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry* 39:379–440.

Hayes, Bruce, Kie Zuraw, Péter Siptár, and Zsuzsa Cziráky Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85:822–863.

Heinz, Jeffrey. 2010. Learning long-distance phonotactics. *Linguistic Inquiry* 41:623–661.

Jardine, Adam. 2015. Learning tiers for long-distance phonotactics. In *Proceedings of the 6th Conference on Generative Approaches to Language Acquisition North America.*

Jardine, Adam, and Jeffrey Heinz. 2016. Learning Tier-based Strictly 2-Local Languages. *Transactions of the Association for Computational Linguistics* 4:87–98.

Kastner, Itamar, and Frans Adriaans. 2017. Linguistic Constraints on Statistical Word Segmentation: The Role of Consonants in Arabic and English. *Cognitive Science* .

Kimper, Wendell. 2011. Competing triggers: Transparency and opacity in vowel harmony. Doctoral dissertation, University of Massachusetts, Amherst, MA.

Kochetov, Alexei, and Milica Radisic. 2009. Latent consonant harmony in Russian: Experimental evidence for agreement by correspondence. In *Proceedings of the 17th meeting of Formal Approaches to Slavic Linguistics*, ed. Maria Babyonyshev, Darya Kavitskaya, and Jodi Reich, 111–130. Ann Arbor: Michigan Slavic Publications.

MacEachern, Margaret. 1997. Laryngeal Cooccurrence Restrictions. Doctoral Dissertation, UCLA, Los Angeles.

Maddieson, Ian. 1990. Shona velarization: complex consonants or complex onsets? *UCLA Working Papers in Linguistics* 74:16–34.

McCarthy, John J. 1986. OCP Effects: Gemination and antigemination. *Linguistic Inquiry* 17:207–263.

McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 43:84–108.

McCarthy, John J. 1989. Linear order in phonological representation. *Linguistic Inquiry* 20:71–99.

McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. In *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, ed. Patricia Keating, 191–233. Cambridge: Cambridge University Press.

Mester, Armin. 1986. Studies in Tier Structure. Doctoral Dissertation, University of Massachusetts, Amherst. Published 1988. Outstanding Dissertations in Linguistics series. New York: Garland.

Mudzingwa, Calisto. 2010. Shona morphophonemics: Repair strategies in Karanga and Zezuru. Doctoral dissertation, University of British Columbia, Vancouver, BC.

Myers, Scott. 1987. Tone and the Structure of Words in Shona. Doctoral Dissertation, University of Massachusetts, Amherst.

Padgett, Jaye. 1991. Stricture in feature geometry. Doctoral Dissertation, University of Massachusetts, Amherst, Amherst, MA.

Rose, Sharon, and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80:475–532.

Siptár, Péter, and Miklós Törkenczy. 2000. *The phonology of Hungarian.* Oxford: Oxford University Press.

Stanton, Juliet. 2017a. Constraints on the distribution of nasal-stop sequences: An argument for contrast. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Stanton, Juliet. 2017b. *Latin –alis/–aris and segmental blocking in dissimilation.*

Stanton, Juliet. to appear. Learnability shapes typology: the case of the midpoint pathology. *Language* .

Suzuki, Keiichiro. 1998. A Typological Investigation of Dissimilation. Doctoral Dissertation, University of Arizona, Tucson, AZ.

Svantesson, Jan-Olof, Anna Tsendina, Anastasia Karlsson, and Vivan Franzén. 2005. The Phonology of Mongolian. *Phonology of the World's Languages* .

Trubetzkoy, N. S. 1939. *Grundzuge der Phonologie.* Prague: Travaux du cercle linguistique de Prague 7.

Van Kampen, Anja, Güliz Parmaksiz, Ruben van de Vijver, and Barbara Höhle. 2008. Metrical and statistical cues for word segmentation: The use of vowel harmony and word stress as cues to word boundaries by 6-and 9-month-old Turkish learners. *Language acquisition and development: Proceedings of GALA 2007* 2007:313–324.

Walker, Rachel. 2001. Round Licensing, Harmony, and Bisyllabic Triggers in Altaic. *Natural Language & Linguistic Theory* 19:827–878.

Walker, Rachel, Dani Byrd, and Fidèle Mpiranya. 2008. An articulatory view of Kinyarwanda coronal harmony. *Phonology* 25:499–535. URL `http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=4292312&fulltextType=RA&fileId=S0952675708001619`.

Wilson, Colin, and Gillian Gallagher. 2018. Constraint complexity in surface-based phonotactics: A case study of South Bolivian Quechua. *Linguistic Inquiry* .

Wilson, Colin, and Marieke Obdeyn. 2009. Simplifying subsidiary theory: Statistical evidence from Arabic, Muna, Shona, and Wargamay. Ms.

Zuraw, Kie. 2002. Aggressive reduplication. *Phonology* 19:395–439.

Zuraw, Kie, and Bruce Hayes. 2017. Intersecting constraint families: an argument for Harmonic Grammar. *Language* 93:497–548.

Zuraw, Kie, and Yu-An Lu. 2009. Diverse repairs for multiple labial consonants. *Natural Language & Linguistic Theory* 27:197–224.