

Phonotactic Restrictions and Morphology in Aymara^{*}

Gillian Gallagher, Maria Gouskova, and Gladys Camacho Rios

Abstract

Nonlocal phonological interactions are often sensitive to morphological domains. Bolivian Aymara restricts the co-occurrence of plain, ejective, and aspirated stops within but not across morphemes. We document these restrictions in a morphologically parsed corpus of Aymara, where they hold as a strong statistical tendency. We further present two experiments with native Aymara speakers. In the first experiment, speakers are asked to repeat nonce words that should be interpreted as monomorphemic. Speakers are more accurate at repeating nonce words that respect the nonlocal phonotactic restrictions than nonce words that violate them. In a second experiment, some nonce words are interpretable as morphologically complex, while others suggest a monomorphemic parse. Speakers show a sensitivity to this difference, and repeat the words more accurately when they can be interpreted as having a morpheme boundary between two consonants that tend to not co-occur inside a morpheme. Finally, we develop a computational model that induces nonlocal representations from the baseline grammar. The model is triggered to posit projections when it notices that certain segments often co-occur when separated by a morpheme boundary but not a single segment. The model generates a full Maximum Entropy phonotactic grammar, which makes distinctions between attested and rare/unattested sequences in a way that closely aligns with the behavior of Aymara speakers.

1 Introduction

One of the challenges for a theory of phonotactics is recognizing that constraints can hold inside morphemes but be lifted at morpheme boundaries (Trubetzkoy, 1939, Chomsky and Halle, 1968, et seq.). This situation is quite common: for example, in English, the cluster [md] is not found inside morphemes, but it is allowed in suffixed verbs such as “hemmed”. Likewise, when it

^{*}For feedback on this and related work, we would like to thank audiences at NYU, AMP 6 at UC San Diego, and NELS 49 at Cornell. We would like to thank Ildikó Emese Szabó for assistance with building the corpus, and Colin Wilson for sharing the code for the gain-based version of the MaxEnt Phonotactic Learner. This research was supported in part by NSF BCS-1724753 to the first two authors.

comes to nonlocal phonological interactions, some languages respect the relevant constraints in any phonological word, but it seems to be equally if not more common for nonlocal phonotactics to apply differently inside vs. across morphemes. These kinds of patterns present an interesting learnability problem: if a learner attends to phonological words only, the relevant constraints may be violated, so how, if at all, do speakers arrive at the knowledge of correct constraints that hold morpheme-internally?

Our paper investigates this question in a study of nonlocal phonological interactions in Bolivian Aymara. Within Aymara morphemes, plain-aspirate, plain-ejective and heterorganic ejective-ejective combinations are described as restricted (see (1)), though these combinations may arise across morpheme boundaries (see (2)):

(1) Aymara laryngeal phonotactics inside morphemes (MacEachern 1997)

<i>ordering restriction on laryngeals:</i>		<i>no co-occurrence of non-identical ejectives:</i>	
a. p ^h utu	‘heat’	*tup ^h u	
b. k ^h apa	‘cartilage’	*kap ^h a	
c. k ^h ask’a	‘acid to the taste’	*t ^h ank’a	
d. t ^h ant’a	‘bread’		

(2) Aymara laryngeal phonotactics across morpheme boundaries (our fieldwork data)

plain-ejective combinations attested across morpheme boundaries

- a. pa^h-t’a-*ɲa* ‘about to choose’
 b. t*ɲ*-t^huki-*ɲa* ‘to color carefully’

plain-aspirate combinations attested across morpheme boundaries

- c. t^ha^hm-t^hapi-*ɲa* ‘to finish chewing’
 d. qa^h-t^hapi-*ɲa* ‘to finish scratching’

ejective-ejective combinations attested across morpheme boundaries

- e. t^hum-t’a-*ɲa* ‘about to drain’
 f. t^hisn-t^huki-*ɲa* ‘to thread carefully’

Impressionistic descriptions of phonological patterns are often made more nuanced by explorations of natural language corpora, experimentation with native speakers, and computational modeling. In this paper, we look at phonological generalizations in Bolivian Aymara through these three lenses.

Our examination of a morphologically parsed web corpus partially confirms traditional descriptions in the literature for the plain-ejective and plain-aspirate restrictions: while there numerous exceptions to the restriction in tautomorphemic combinations, there are far more exceptions heteromorphemically than tautomorphemically. Ejective-ejective combinations, however, are infrequent regardless of morphological context.

Despite the exceptions in the lexicon and the overall infrequency of ejective-ejective com-

binations, two experiments support the synchronic status of restrictions on plain-ejective and ejective-ejective combinations. Native Aymara speakers make more repetition errors on nonce words that violate the putative restrictions than on control words, and speakers make fewer errors on nonce words when the interacting stops may be interpreted as belonging to different morphemes than when they must be parsed as tautomorphemic.

After establishing the corpus and behavioral evidence for the restrictions, we present a computational model that learns the morphologically sensitive, nonlocal phonotactic restrictions from our corpus. The modeling work shows that while certain aspects of the phonotactic restrictions are observable in an unparsed data set (cf. Martin 2007), training on a parsed corpus with morpheme boundaries is necessary to capture the full range of patterns in our experiments and the descriptive literature.

The modeling work in this paper expands on the model developed in Gouskova and Gallagher (2017). There, we proposed a method for inductively learning nonlocal projections that capitalizes on the observation that nonlocal interactions can be observed in local phonotactics: if X and Y cannot co-occur at longer distances inside a word, they usually cannot be separated by a single segment, either (Suzuki 1998). Our learner induces nonlocal projections by attending to the properties of the language’s segment-level phonotactics. In languages with nonlocal phonological interactions, segments within a certain natural class are restricted from co-occurring across an arbitrary amount of intervening material: e.g., in Quechua pairs of ejectives may not co-occur across an intervening vowel *[k’ap’i], an intervening vowel and consonant *[k’amp’i] or across more material *[k’amip’a] (Gallagher 2016). The arbitrary nature of the intervening segmental material has supported analyses of these patterns that reference an autosegmental tier or *projection*¹ where only the interacting segments are visible. For the Quechua case, this would mean that there is one level of representation in which all segments are visible to the grammar, and another level of representation in which only ejectives are visible; it is on this ‘ejective projection’ that the co-occurrence restriction can be stated as a simple bigram *[+cg][+cg].

Hayes and Wilson’s (2008) inductive phonotactic learner allows the analyst to define nonlocal projections so the model can learn nonlocal phonology. We propose that nonlocal projections can be learned inductively by analyzing the constraints in a baseline grammar without projections. In languages with nonlocal phonology, the baseline grammar will sometimes include trigram constraints of the form *A_any_segment_B. A trigram of this sort is a clue to the learner that natural classes A and B interact nonlocally, and that the nature of the intervening material is

¹The difference between ‘tier’ and ‘projection’ has to do with representational assumptions: the term *tier* has historically meant a level of a structured autosegmental representation, whereas a *projection* is merely a representation that includes all and only the members of some class, e.g., all the vowels in a word on a [+syllabic] projection. We adopt the latter term following Hayes and Wilson (2008). See Clements (1976), Goldsmith (1976), McCarthy (1979), Archangeli (1985), McCarthy (1989) and many others.

irrelevant. Our original model builds projections based on these trigram constraints, and in this paper we expand the procedure to also induce projections from morpheme boundary trigrams: A__[-mb]_B, where [-mb] is the class of all segments but not the morpheme boundary symbol. Intuitively, these constraints will arise in a language where the segments A and B cannot co-occur inside a morpheme (*A-any_{segment}-B), but occur with some frequency at morpheme boundaries (✓A+B). Constraints of this form indicate that natural classes A and B interact nonlocally, but strictly tautomorphemically. The simulations reported in this paper show that the morphologically sensitive, nonlocal restrictions in Aymara are observable as morpheme boundary trigrams in a parsed corpus, despite the presence of exceptions.

The paper is structured as follows. Section 2 summarizes the laryngeal constraints that hold of Aymara words—we cover the descriptive generalizations in the literature on the language, and present our own study of a web corpus of Aymara. Section 3 presents two experimental studies with Aymara speakers, which test their knowledge of nonlocal phonotactics that hold of morphologically simple words as opposed to complex ones in a nonce word repetition experiment. Section 4 presents our computational model and a simulation that induces nonlocal projections from the web corpus described in section 2. Section 5 offers some general discussion, and section 6 concludes the paper.

2 Laryngeal restrictions in Aymara

2.1 Background

The consonant inventory of Aymara contains fifteen stops, exhibiting a three-way laryngeal contrast between plain (voiceless unaspirated), ejective and aspirate at five places of articulation. The full inventory is shown in Table 1 (MacEachern 1997, Hardman 2001).

	labial	dental	postalveolar	velar	uvular	glottal
plain	p	t	tʃ	k	q	
ejective	p'	t'	tʃ'	k'	q'	
aspirate	p ^h	t ^h	tʃ ^h	k ^h	q ^h	
fricative		s	ʃ		χ	h
nasal	m	n	ɲ			
liquid		l r	ʎ			
glide	w		j			

Table 1: Consonant inventory of Aymara

The distribution of ejective and aspirate stops is restricted within roots, which are primarily CV(C)CV. As shown in (3a), ejectives and aspirates may appear in either initial or medial position.

Both ejectives and aspirates are rare in roots with an initial plain stop, however (see (3b)). Pairs of heterorganic ejectives are also rare, though forms with identical ejectives are attested (see (3c)). Other combinations of ejectives and aspirates are attested (see (3d)). Examples are from De Lucca (1987), and these and other patterns are also discussed in detail in MacEachern (1997) and Bennett (2013).

(3) Aymara ejective and aspirate distribution

- | | | | | |
|----|----------|----------------|---------|-----------|
| a. | tʰaku | ‘coarse’ | k’atʃa | ‘voice’ |
| | laqʰa | ‘darkness’ | hajp’u | ‘evening’ |
| b. | *paqʰa | | *kajp’u | |
| c. | p’ap’i | ‘roasted fish’ | *k’ap’i | |
| | pʰuʌtʃ’u | ‘bag’ | tʰakʰi | ‘road’ |
| | k’ampʰi | ‘tip over’ | | |

The three laryngeal combinations that are underattested in roots—plain-ejective, plain-aspirate and ejective-ejective—are attested in words. These combinations arise when suffixes with an ejective or aspirate consonant combine with roots with a plain or ejective stop. Some examples are given in (4), from work with a native speaker consultant in El Alto, Bolivia. These examples involve three verbal suffixes, [-t’a] ‘about to’, [-tʃ’uki] ‘carefully, continuously’, and [-tʰapi] ‘finish’. All three of these suffixes trigger deletion of the root final vowel.

(4) Ejectives and aspirates in morphologically complex words of Aymara

- | | | | |
|----|-------------------|---------------------|--|
| a. | plain-ejective | | |
| | paʌ+t’a+ɲa | ‘about to choose’ | tɪɲ+tʃ’uki+ɲa ‘to color carefully’ |
| | taw+t’a+ɲa | ‘about to row’ | pump+tʃ’uki+ɲa ‘to mix carefully’ |
| b. | plain-aspirate | | |
| | tʃaʌm+tʰapi+ɲa | ‘to finish chewing’ | qaq+tʰapi+ɲa ‘to finish scratching’ |
| c. | ejective-ejective | | |
| | tʃ’um+t’a+ɲa | ‘about to drain’ | t’isn+tʃ’uki+ɲa ‘to thread carefully’ |
| | q’ex+t’a+ɲa | ‘about to whip’ | k’uɲ+tʃ’uki+ɲa ‘to bend over continuously’ |

There are several exceptions to the restrictions on tautomorphic ejective-ejective, plain-ejective and plain-aspirate combinations. These are given in (5). Some of these exceptions occur at the level of a trigram on the linear string - important for our model - while others occur across more intervening material and would be noticeable only when looking at a nonlocal projection. Additionally, there are four combinations of plain-ejective and two combinations of plain-aspirate that occur in clusters. These are reported by our consultant to be monomorphemic forms, though

Hardman claims that root-internal stop codas are not found. The morphological structure of these forms is thus in question. An additional observation is that several of the exceptions end in the sequence [t'a], just like the productive suffix.

(5) Exceptions to restrictions

a. plain-aspirate

tap^hijala 'earthen wall' q^hatfq^ha 'rough to the touch'

kaw^ha 'where' t^hapt^ha 'mediocre'

b. plain-ejective

pist'a 'scarcity' λupt'a 'bribery' lupt'a 'when it is very hot'

qat'i 'type of potato' loqt'a 'scope' ukt'a 'height'

c. ejective-ejective

q'ewt'a 'curve, angle'

2.2 Descriptive lexical statistics

To assess the statistical evidence available to an inductive learner trying to acquire these restrictions, we looked at the observed combinations of all three series of stops. Our data set is a morphologically segmented word list. This list was compiled by taking an unsegmented web corpus of 88,728 forms, collected from 438 webpages by An Crúbadán (<http://crubadan.org/>).² The corpus was converted to lowercase and cleaned to remove numbers, non-alphanumeric characters and English or Spanish forms. Forms were also removed if they contained stray apostrophes, hyphens or other typos that we couldn't interpret. This list was then crossed with a list of 1846 roots, derived from the De Lucca (1987) dictionary with the help of a native speaker consultant, and a list of 50 suffixes and their allomorphs from the Hardman (2001) grammar and the De Lucca (1987) dictionary. There were 46,164 forms that were divisible into a known root and known suffixes, and these forms comprise the corpus we use in this paper.

Before looking at combinations of stops directly, we report on the distribution of stops by position in our word corpus, comparing the number of each class of stops in root initial position, root medial position, and in a suffix. Table 2 gives the raw counts on the left (e.g., there are 9,113 plain stops which are in root-initial position in the word corpus), as well as the probability of a stop from the given class in the given position (e.g., 20% of our 1,846 roots begin with a plain stop, 7% with an aspirate and 7% with an ejective, the remaining 66% of roots begin with a vowel, fricative or sonorant consonant). These numbers show that plain stops are frequent in both roots

²The corpus posted on the website is cut off at 50,000 forms. The full version of the corpus was obtained via personal communication with the developers.

and suffixes, while aspirates and ejectives are both much more frequent in roots than in suffixes.

	root initial	root medial	suffix
plain	9,113 (0.20)	17,441 (0.12)	65,148 (0.23)
aspirate	3,007 (0.07)	2,871 (0.02)	817 (<0.01)
ejective	3,037 (0.07)	2,622 (0.02)	1,901 (<0.01)

Table 2: Observed occurrences and probability of stops in three positions.

Tables 3 and 4 report the observed counts for stop combinations in tautomorphemic and heteromorphemic strings, on both the baseline and a nonlocal projection containing only stops. For tautomorphemic sequences, we looked at the co-occurrence of nonadjacent stops in a baseline trigram—that is, of a C_1XC_2 string, where X can be any segment except the morpheme boundary symbol—and for adjacent bigrams on the stop projection. For heteromorphemic sequences, we looked at trigrams where the medial gram was the morpheme boundary on the baseline and the stop projection.

Table 3 shows that, within a baseline trigram, the restricted combinations are unattested tautomorphemically. Ejective-ejective combinations are also nearly unattested in heteromorphemic contexts, but plain-aspirate and plain-ejective combinations are more frequent across a morpheme boundary. The bottom portion of the table shows that other combinations of stops are either frequent in both heteromorphemic and tautomorphemic contexts, or are more frequent tautomorphemically than heteromorphemically (note that while ejectives and aspirates may occur in suffixes, the numbers here reflect the rarity of such suffixes in the corpus as a whole).

	tautomorphemic	<i>example</i>	heteromorphemic	<i>example</i>
plain-aspirate	0	... <i>pat</i> ^h <i>a</i> ...	149	... <i>lip</i> + <i>t</i> ^h <i>a</i> ...
plain-ejective	0	... <i>pat</i> ' <i>a</i> ...	659	... <i>lip</i> + <i>t</i> ' <i>a</i> ...
ejective-ejective	0	... <i>p</i> ' <i>at</i> ' <i>a</i> ...	1	... <i>lip</i> '+' <i>t</i> ' <i>a</i> ...
plain-plain	3532	... <i>pata</i> ...	3673	... <i>lip</i> + <i>ta</i> ...
aspirate-plain	683	... <i>p</i> ^h <i>ata</i> ...	30	... <i>lip</i> ^h + <i>ta</i> ...
ejective-plain	765	... <i>p</i> ' <i>ata</i> ...	46	... <i>lip</i> '+' <i>ta</i> ...
aspirate-aspirate	613	... <i>p</i> ^h <i>at</i> ^h <i>a</i> ...	0	... <i>lip</i> ^h + <i>t</i> ^h <i>a</i> ...
ejective-aspirate	466	... <i>p</i> ' <i>at</i> ^h <i>a</i> ...	4	... <i>lip</i> '+' <i>t</i> ^h <i>a</i> ...
aspirate-ejective	38	... <i>p</i> ^h <i>at</i> ' <i>a</i> ...	2	... <i>lip</i> ^h + <i>t</i> ' <i>a</i> ...

Table 3: Corpus counts for tautomorphemic and heteromorphemic stop combinations in a baseline trigram. Examples are schematic and ellipses represent any preceding or following material.

Table 4 gives the counts on a stop projection. Here, both [p'at'] and [p'ant'] would count as ejective-ejective combinations. Plain-aspirate and plain-ejective combinations are still much more frequent in heteromorphemic than tautomorphemic contexts, though there are substantially

more exceptions tautomorphemically than are observable in a baseline trigram. Ejective-ejective combinations are again rare in both contexts. The bottom portion of the table shows that most combinations (plain-plain, aspirate-plain, ejective-plain) are well attested in both morphological contexts, while aspirate-ejective combinations are somewhat rare in both contexts. Aspirate-aspirate and ejective-aspirate combinations are both more frequent in tautomorphemic combinations, again due to the general rarity of aspirates in suffixes.

	tautomorphemic	<i>example</i>	heteromorphemic	<i>example</i>
plain-aspirate	61	... <i>past^ha</i> ...	261	... <i>pas+t^ha</i> ...
plain-ejective	68	... <i>past'a</i> ...	668	... <i>pas+t'a</i> ...
ejective-ejective	4	... <i>p'ast'a</i> ...	7	... <i>p'as+t'a</i> ...
plain-plain	5389	... <i>pasta</i> ...	23519	... <i>pas+ta</i> ...
aspirate-plain	943	... <i>p^hasta</i> ...	1262	... <i>p^has+ta</i> ...
ejective-plain	906	... <i>p'asta</i> ...	1781	... <i>p'as+ta</i> ...
aspirate-aspirate	712	... <i>p^hast^ha</i> ...	1	... <i>p^has+t^ha</i> ...
ejective-aspirate	476	... <i>p'ast^ha</i> ...	22	... <i>p'as+t^ha</i> ...
aspirate-ejective	38	... <i>p^hast^ha</i> ...	29	... <i>p^has+t^ha</i> ...

Table 4: Corpus counts for tautomorphemic and heteromorphemic stop combinations on a stop projection. Examples represent cases where stops are adjacent on a projection but non-adjacent on the baseline.

The numbers here show that the restrictions on plain-ejective and plain-aspirate combinations in descriptive grammars are supported in counts over a word corpus, at both the baseline trigram level and on a nonlocal projection including only stops. The restrictions on ejective-ejective combinations are more difficult to assess, due to the rarity of heteromorphemic combinations (though our consultant work shows that these are possible, if not frequent in the corpus). The other six combinations of stops are reported to be licit in all morphologically contexts, and this appears to be essentially true in our corpus as well, though certain combinations are unattested or nearly unattested (See §4.4.4 for discussion of other restricted combinations).

The experiments presented in the next section look at how native speakers of Aymara treat forms with combinations of plain-ejective and ejective-ejective combinations, in both heteromorphemic and tautomorphemic contexts. We then present the results of our computational model in §4, showing how the counts presented above are reflected in an inductive phonotactic grammar.

3 Experimental work

The two experiments reported below provide behavioral evidence that speakers of Aymara have learned the laryngeal restrictions on ejectives, and that treatment of these phonotactic structures

is sensitive to morphological structure.

3.1 Experiment 1: sensitivity to ejective co-occurrence and ordering restrictions

Experiment 1 presents Aymara speakers with simple disyllabic forms, which could be interpreted as pseudo nouns, that contain plain-ejective or ejective-ejective combinations. Speakers' errors in repeating such forms are evaluated to assess whether speakers have internalized phonotactic restrictions on these combinations.

3.1.1 Participants

The participants were 21 native Aymara speakers, all Spanish bilinguals. All were college educated and resided in El Alto, Bolivia, and most were students at Universidad Pública de El Alto. There were seven male and fourteen female participants, aged 19-30. Fifteen participants reported that they had been speaking Aymara since birth, five learned Aymara between the ages of 4 and 7, and one at 12.

3.1.2 Methods

Stimuli The stimuli were disyllabic C_1VC_2V nonce forms. Control items had a phonotactically legal ejective in C_2 and a fricative or sonorant in C_1 . Ejective-ejective items contained a putative phonotactic violation by having heterorganic ejectives in C_1 and C_2 and plain-ejective items contained a putatively restricted combination of a plain stop in C_1 and an ejective in C_2 . There were fifteen items of each type, and an additional fifteen, phonotactically legal fillers with a plain stop in C_2 , for a total of 60 items. The complete list of stimuli is shown in Table 5.

control			ejective-ejective			plain-ejective			filler		
lap'a	saq'o	nut'a	k'it'a	p'ik'a	k'ap'u	tip'i	tuk'i	kip'a	t'apu	kupa	lipu
jup'a	moq'o	yap'i	p'it'a	t'oq'e	k'ut'a	kut'a	toq'e	kap'i	k'api	tipu	napu
juk'u	lip'u	lik'a	q'ap'i	t'ap'u	k'up'i	pit'a	kup'a	tip'a	k'ati	kipi	natu
juk'a	nap'u	luk'a	k'ip'a	k'ip'i	p'uk'a	kip'u	kap'a	puk'i	k'upi	kapu	japi
seq'a	nat'u	maq'o	q'at'a	q'op'i	t'aq'o	qat'i	tip'u	taq'e	p'uka	puki	luka

Table 5: Stimuli for experiment 1

The stimuli were made from recordings of a native Aymara speaking consultant reading phonotactically legal nonce words. The stimuli were made by splicing together C_1V and C_2V

during the closure of the second stop, e.g., [lap'a] was made by splicing [lapa] and [map'a] together during the labial closure. All stimuli were normalized for amplitude, but were otherwise unmodified.

Procedure Participants were seated in front of a laptop computer wearing AudioTechnica noise cancelling headphones. The stimuli were presented using PsyScope (<http://psy.ck.sissa.it/>). On each trial, the audio stimulus was played once and participants were asked to repeat what they heard as precisely as possible. Participants were told that the words they would hear were not real words of Aymara, though they would contain sounds familiar from Aymara. Once participants had repeated the item, they pressed any key on the keyboard to move on to the next trial. No orthographic representation of the stimuli was given.

Analysis The audio recordings of participants' responses were transcribed, and coded for accuracy and type of error, if any. Errors on ejective-ejective and plain-ejective items were then further classified as repairs or non-repairs, depending on whether they removed the putative phonotactic violation or not.

3.1.3 Results

Accuracy Overall accuracy differed between control, ejective-ejective and plain-ejective items, as shown in Figure 1. Accuracy on control items was very high (97%), while accuracy on ejective-ejective and plain-ejective items was lower, consistent with a difference in phonotactic legality. Items with an ejective-ejective combination were repeated accurately more often than items with a plain-ejective combination (50% vs. 32%).

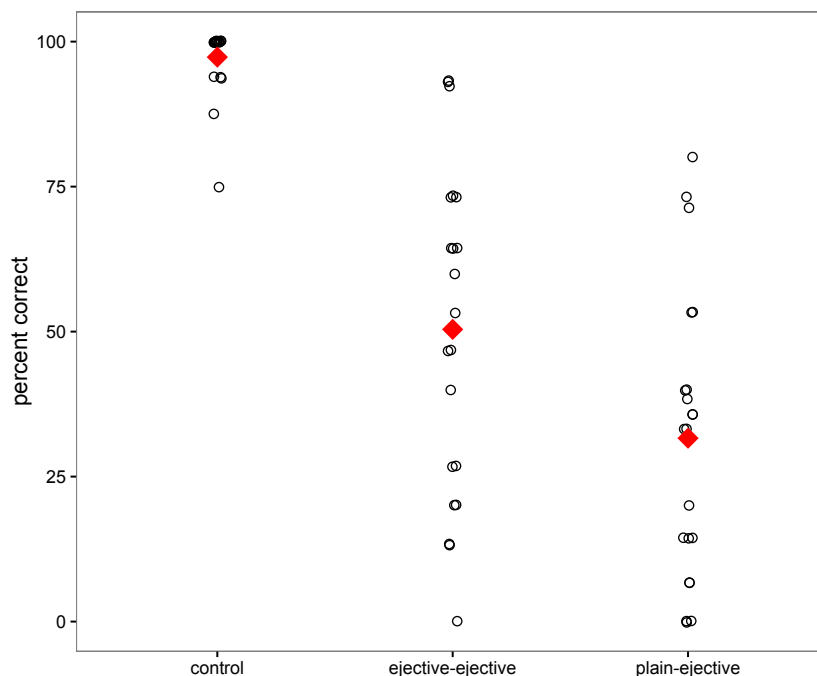


Figure 1: Accuracy on control, ejective-ejective and plain-ejective forms in Experiment 1. Open circles indicate an individual participant’s performance; red diamonds show the mean accuracy across all participants.

All trials were coded for accuracy (correct or incorrect), and a binomial, linear mixed model was then fit with accuracy as the dependent variable, a ternary predictor of type, a random effect of type by subject and a random slope by subject. Ejective-ejective was set as the baseline to which the other two factor levels, plain-ejective and control, were compared. The model was fit using the `lmer` function in the `lme4` package (Bates et al. 2014) for R (R Development Core Team 2018, <https://www.r-project.org/>). Both comparisons were significant. Accuracy on control is significantly higher than accuracy on ejective-ejective ($\beta = 4.81$, $SE = 1.02$, $z = 4.72$, $p < 0.0001$), and accuracy on plain-ejective stimuli is significantly lower than on ejective-ejective stimuli ($\beta = -1.03$, $SE = 0.44$, $z = -2.31$, $p = 0.02$).

To allow comparison with Experiment 2 below, the results of Experiment 1 were also analyzed for an effect of place of articulation. Plain-ejective and ejective-ejective trials were coded for whether the medial ejective was dental or not (labial, velar or uvular), and a model was fit to accuracy (correct or incorrect) with type (plain-ejective or ejective-ejective), place (dental or not) and their interaction as predictors. The model had a random intercept by subject and a random slope for place (a model with random slopes for place and type failed to converge). The interaction between type and place was not significant, so it was removed from the model. In the model without the interaction, the main effect of type was again significant (accuracy on plain-ejective

stimuli is lower than on ejective-ejective stimuli, $\beta = -0.90$, $SE = 0.19$, $z = -4.83$, $p < 0.0001$), and the model also found a main effect of place. Accuracy on forms with a dental ejective in C_2 is higher than on forms with a non-dental ($\beta = 0.67$, $SE = 0.29$, $z = -2.30$, $p = 0.02$).

Errors The frequency of different errors on ejective-ejective and plain-ejective stimuli are summarized in Table 6 and Table 7, distinguishing between repair and non-repair errors.

	error	example	% of responses	(n)
repair	C_2 de-ejectivization	k'ap'u → k'apu	39%	(121)
	C_1 and C_2 de-ejectivization	k'ap'u → kapu	2%	(6)
	C_1 deletion	k'ap'u → ap'u	0.5%	(1)
	C_2 aspiration	k'ap'u → k'aphu	1%	(4)
	<i>total</i>		42.5%	
non-repair	C_1 de-ejectivization	k'ap'u → kap'u	4%	(12)
	labial → dental	k'ap'u → k'at'u	3.5%	(11)
	<i>total</i>		7.5%	

Table 6: Errors on ejective-ejective stimuli

	error	example	% of responses	(n)
repair	C_2 de-ejectivization	kap'u → kapu	12%	(37)
	ejective reassociation	kap'u → k'apu	28%	(86)
	C_1 change	kap'u → hap'u	5%	(16)
	<i>total</i>		45%	
non-repair	ejective doubling	kap'u → k'ap'u	22%	(66)
	labial → dental	kap'u → kat'u	1%	(3)
	C_2 aspiration	kap'u → kaphu	1%	(2)
	<i>total</i>		24%	

Table 7: Errors on plain-ejective stimuli

Errors on both types of stimuli remove the putative phonotactic violation more often than not, though ejective doubling errors are quite common on plain-ejective stimuli. Looking at the distribution of errors, we can see that the difference in accuracy between ejective-ejective and plain-ejective stimuli does not stem from a difference in how often these structures are repaired; repair rates for the two stimulus types are comparable. Instead, the lower accuracy on plain-ejective forms overall is driven by the greater rate of non-repair errors see Gallagher (2016) for further discussion of this type of error. The relevance of errors that map a labial ejective to a dental will be discussed in conjunction with the results of Experiment 2 below.

3.1.4 Discussion

The results of Experiment 1 support the status of both ejective-ejective and plain-ejective combinations as synchronically restricted in Aymara. Forms that violate these restrictions are repeated significantly less accurately than phonotactically legal controls.

There is also an effect of place of articulation, with higher accuracy on forms with medial dental ejectives. As described above, Aymara has a productive suffix [t'a] that may result in plain-ejective or ejective-ejective combinations at the word level. While the stimuli in Experiment 1 had the shape of bare roots—as opposed to the morphologically complex forms in Experiment 2—the greater accuracy on forms with a dental ejective reflects the likelihood of a dental in C_2 when plain-ejective and ejective-ejective combinations occur at the word level. The independent roles of place of articulation and morphological structure will be discussed further below, by comparing the results of Experiment 1 and Experiment 2.

3.2 Experiment 2: laryngeal restrictions and morphological structure

The goal of Experiment 2 is to test whether speakers' errors on ejective-ejective and plain-ejective combinations are influenced by morphological structure, and whether these restrictions hold across more than a single intervening vowel. Experiment 2 presents participants with the same kinds of phonotactically illegal structures - ejective-ejective and plain-ejective pairs - as in Experiment 1, but in Experiment 2 the nonce words are pseudo verbs ending in the infinitival suffix [-pa]. The experiment compares performance on stimuli where the illegal ejective in C_2 must be interpreted as part of the root vs. stimuli where the illegal ejective in C_2 may be interpreted as part of a productive suffix [t'a]. Additionally, the pseudo roots in Experiment 2 all contain a coda consonant, so the interacting consonants are separated by a VC sequence as opposed to the single V in Experiment 1.

3.2.1 Participants

The participants were 20 of the participants from Experiment 1 (data from one participant were accidentally not recorded). Participants were balanced as to whether they completed Experiment 1 or Experiment 2 first.

3.2.2 Methods

Stimuli The stimuli were trisyllabic pseudo verbs, all of which ended in the infinitival suffix [-pa]. The pseudo verb stem was $C_1V_1CC_2V_2$, where C_1 was either a plain stop or an ejective and C_2V_2 was either [p'a] or [t'a]. Forms with [t'a] were plausibly polymorphemic, while forms with

[p'a] were not, since [t'a] is a productive verbal suffix and there is no suffix [-p'a]. All forms had a coda consonant in the first syllable, because in real words the suffix [-t'a] triggers deletion of a root vowel and thus forms a cluster with the final root consonant.

The test forms just described fell into one of four categories, based on the laryngeal restriction that was violated and the place of articulation of C₂ (as a stand-in for implied morphological complexity): plain-ejective-labial, plain-ejective-dental, ejective-ejective-labial, ejective-ejective-dental. There were ten tokens in each test category and 40 filler items, which had a plain stop in C₂ and a plain stop, fricative or sonorant in C₁, for a total of 80 items. The test items are given in Table 8.

ej-ej-labial	ej-ej-dental	pl-ej-labial	pl-ej-dental
k'asp'a+ɲa	k'as+t'a+ɲa	kasp'a+ɲa	kas+t'a+ɲa
k'isp'a+ɲa	k'is+t'a+ɲa	kisp'a+ɲa	kis+t'a+ɲa
k'aɬp'a+ɲa	k'aɬ+t'a+ɲa	kaɬp'a+ɲa	kaɬ+t'a+ɲa
k'uɬp'a+ɲa	k'uɬ+t'a+ɲa	kuɬp'a+ɲa	kuɬ+t'a+ɲa
tʃ'imp'a+ɲa	tʃ'in+t'a+ɲa	tʃimp'a+ɲa	tʃin+t'a+ɲa
tʃ'amp'a+ɲa	tʃ'an+t'a+ɲa	tʃamp'a+ɲa	tʃan+t'a+ɲa
tʃ'uɬp'a+ɲa	tʃ'uɬ+t'a+ɲa	tʃuɬp'a+ɲa	tʃuɬ+t'a+ɲa
tʃ'iɬp'a+ɲa	tʃ'iɬ+t'a+ɲa	tʃiɬp'a+ɲa	tʃiɬ+t'a+ɲa
q'asp'a+ɲa	q'as+t'a+ɲa	qasp'a+ɲa	qas+t'a+ɲa
q'oɬp'a+ɲa	q'oɬ+t'a+ɲa	qoɬp'a+ɲa	qoɬ+t'a+ɲa

Table 8: Stimuli for Experiment 2 (not including fillers)

The stimuli were made from recordings of a native Aymara speaker producing phonotactically legal nonce words, using the same splicing method and normalization as described for Experiment 1.

Procedure The procedure was identical to Experiment 1.

Analysis The analysis was identical to Experiment 1.

3.2.3 Results

Results from three participants were removed from further analysis because they had a low accuracy rate on filler items (15%, 32% and 64%), showing that they struggled with the task as a whole. The following discussion reflects the results of the remaining 17 participants.

Accuracy Accurate repetition of forms with [t'a], those that are plausibly polymorphemic, was higher than those with [p'a], where the plain-ejective or ejective-ejective combination must be

interpreted as monomorphemic (35% vs. 74%). This distinction held for both ejective-ejective and plain-ejective combinations.

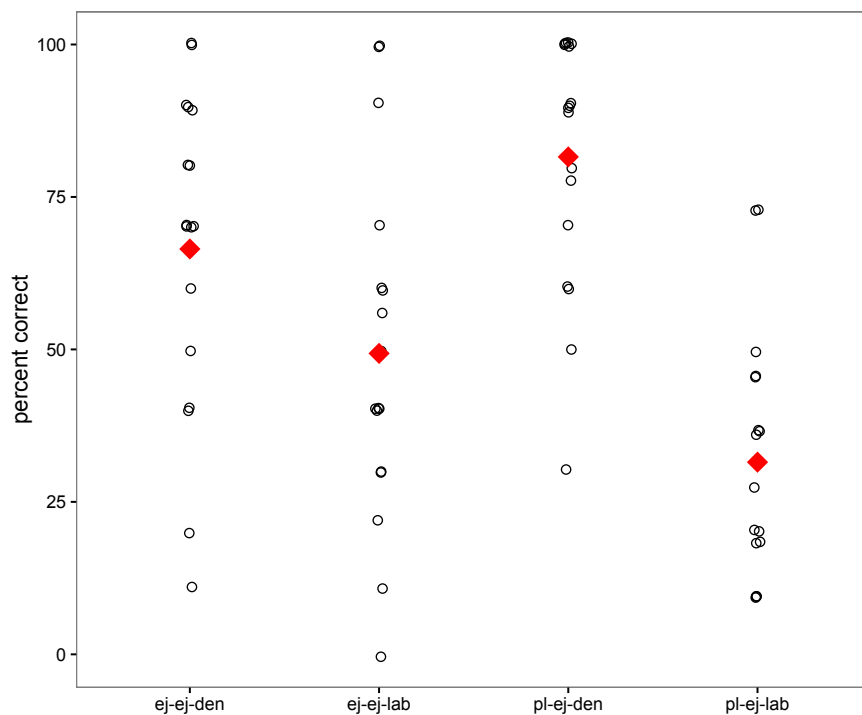


Figure 2: Accuracy in Experiment 2. Open circles indicate an individual participant's performance; red diamonds show the mean accuracy across all participants.

A binomial, linear mixed model was fit to accuracy with predictors of place of articulation (labial or dental), violation type (ejective-ejective or plain-ejective), and their interaction, along with by-participant random slopes for place and type (a model with the interaction as a random by-participant slope failed to converge) and a random intercept for participant. The model finds a main effect of place, with lower accuracy on labial forms than dental forms ($\beta = -1.37$, $SE = 0.31$, $t = -4.39$, $p < 0.0001$). The model also revealed a significant interaction between place and violation type ($\beta = -1.52$, $SE = 0.62$, $t = -2.46$, $p = 0.02$). While overall accuracy only marginally differs between ejective-ejective and plain-ejective violations ($\beta = 1.11$, $SE = 0.59$, $t = 1.89$, $p = 0.06$), the direction of the effect differs depending on place. For labials, accuracy on ejective-ejective is slightly higher than on plain-ejective (38.5% vs. 32%), while for dentals, plain-ejective accuracy is higher than ejective-ejective accuracy (82% vs. 67%).

Errors In Experiment 2, a high number of errors involved changing a labial ejective to a dental ejective, thereby repairing the phonotactic violation by separating the combining stops with a morpheme boundary. For example, in [k'asp'ana], the pair of ejectives must be interpreted as

co-occurring within a root ($*[k'asp'a+ɲa]$), while in $[k'ast'ɲa]$ the pair of ejectives may be interpreted as co-occurring across a morpheme boundary ($*[k'ast'a+ɲa]$ and $[k'as+t'a+ɲa]$ are both possible parses). The frequency of different errors is summarized in Tables 9 and 10. In each table, errors in the top section remove the phonotactic violation entirely. In the second section, errors change the plausible morphological structure of the pseudo-verb by changing a labial to a dental, and in the third section errors do not repair the violation. Virtually all errors are much more frequent for labial forms than dental forms, and place errors are only attested for labial forms.

	error	example	labial		dental	
			% resp.	(n)	% resp.	(n)
lar. repairs	C_2 de-ejectivization	$k'asp'ɲa \rightarrow k'aspɲa$	18.5%	(31)	7%	(12)
	C_1 and C_2 de-ejectivization	$k'asp'ɲa \rightarrow kaspɲa$	4%	(6)	1%	(2)
	C_1 change	$k'asp'ɲa \rightarrow asp'ɲa$	0.5%	(1)	2%	(3)
place repairs	C_1 de-ejectivization & place change	$k'asp'ɲa \rightarrow kast'ɲa$	12%	(20)	0%	(0)
	place change	$k'asp'ɲa \rightarrow k'ast'ɲa$	18.5%	(31)	0%	(0)
non-repairs	C_1 de-ejectivization	$k'asp'ɲa \rightarrow kasp'ɲa$	7%	(12)	23%	(39)
	total		60.5%		33%	

Table 9: Errors on co-occurrence stimuli. Top: errors that remove the phonotactic violation, middle: errors that change place and morphological structure, bottom: non-repair errors. Percentages indicate the total rate of errors out of all responses (e.g., 60.5% labial stimuli were produced with errors, and 38.5% were produced without errors).

	error	example	labial		dental	
			% resp.	(n)	% resp.	(n)
lar. repairs	C_2 de-ejectivization	$kasp'ɲa \rightarrow kaspɲa$	15%	(28)	10%	(17)
	ejective reassociation	$kasp'ɲa \rightarrow k'aspɲa$	6%	(11)	0.5%	(1)
	C_1 change	$kasp'ɲa \rightarrow asp'ɲa$	0%	(0)	0.5%	(2)
place repairs	place change	$kasp'ɲa \rightarrow kast'ɲa$	40%	(74)	0%	(0)
	ejective doubling & place change	$kasp'ɲa \rightarrow k'ast'ɲa$	2%	(4)	0%	(0)
non-repairs	ejective doubling	$kasp'ɲa \rightarrow k'asp'ɲa$	5%	(9)	7%	(11)
	total		68%		18%	

Table 10: Errors on ordering stimuli. Top: errors that remove the phonotactic violation, middle: errors that change place and morphological structure, bottom: non-repair errors. Percentages indicate the total rate of errors out of all responses (e.g., 18% of dental stimuli were produced with errors, and 82% were produced without errors)

3.2.4 Discussion

Experiment 2 shows that speakers' treatment of plain-ejective and ejective-ejective combinations is sensitive to the inferred morphological structure of the forms in which they occur. When these

consonant combinations must be interpreted as being tautomorphemic there is a higher error rate than when the combination may be interpreted as heteromorphemic.

3.3 Summary and comparison of Experiments 1 and 2

Together, Experiments 1 and 2 show that Aymara speakers have learned morphologically sensitive restrictions on ejective-ejective and plain-ejective combinations. Comparison between the two experiments supports a role for both morphological structure and phonetic category in the place of articulation effect found in both studies.

Error rates on forms with dental ejectives were lower than error rates on forms with non-dental ejectives in both experiments. This effect is at least partly an effect of phonetic category: dental ejectives are much more common as the second consonant in a plain-ejective or ejective-ejective sequence than other places of articulation (97% of such combinations in our corpus have [tʰ] as C₂), because of the frequency of the suffix [-tʰa]. We can thus conclude that speakers are sensitive to the different distribution of [tʰ] compared to other ejectives.

To see if there is also a contribution of morphological structure, we need to consider the increased accuracy on dental ejective forms between Experiments 1 and 2. In Experiment 1, all forms, including those with dental ejectives, have a monomorphemic structure. Some of the dental ejective stimuli in Experiment 1 cannot be decomposed into suffixes (e.g., [qatʰi])—there is no suffix [tʰi]). Stimuli with [tʰa], such as [katʰa] and [nutʰa], are unlikely to be analyzed as root-suffix because the suffix [tʰa] always attaches to bases that are at least CVC (e.g., [taw-tʰa-ɲa] ‘about to row’).³ The other reason for doubting that the dental ejective stimuli in Experiment 1 are morphologically decomposed is that [tʰa] is an aspectual suffix that is usually followed by tense and person marking. There are only six words in the corpus where [-tʰa] is the last and only suffix, and none of them have the shape *C V tʰ a*.

On the other hand, in Experiment 2, forms with a dental ejective have a polymorphemic structure, since they contain the two verbal suffixes [tʰa] and [ɲa]. To test for an effect of morphological structure, responses to ejective-ejective and plain-ejective stimuli from the two experiments were pooled and a binomial linear mixed model was fitted. The dependent variable was accuracy, and the independent variables were experiment (Experiment 1 or Experiment 2) and place (dental or not). There was a random intercept for participant and a random by-participant slope for place (a model with a random slope for experiment failed to converge). The model found a significant interaction between place and experiment ($\beta = 1.14$, SE = 0.28, $t = 4.15$, $p < 0.0001$), revealing that the effect of place differs between the two experiments. The difference between accuracy on dental and non-dental forms is larger in Experiment 2 (74% vs. 35%, a 39 point difference),

³There are several roots that do not undergo apocope when suffixed with [-tʰa] in our corpus: [hawɪ], [ana], [qoɫa], and [wajka].

where stimuli have a polymorphemic structure, than in Experiment 1 (52% vs. 37%, a 15 point difference) where stimuli have a monomorphemic structure.

The types of errors between the two experiments also differ, and further show the importance of morphological structure to repetition accuracy. While place of articulation errors are quite rare in Experiment 1, these errors are very frequent in Experiment 2. Errors that map a labial ejective to a dental make up just 4.5% of errors in Experiment 1 but 73% percent in Experiment 2. Participants' responses in Experiment 2 are thus tracking the polymorphemic, pseudo-verb structure of the stimuli, skewing responses to create a phonotactically legal form by changing place of articulation and thus morphological structure.

In sum, the experiments here provide behavioral evidence that the laryngeal restrictions on ejectives and their interaction with morphology are part of speakers' synchronic grammars.

4 Learning simulations

Having presented the corpus and behavioral evidence for the restrictions in Aymara, we move on to modeling the learning of these patterns from our corpus. Our model starts with a parsed corpus of word forms, notices the need for nonlocal projections, and induces a set of constraints that capture the local and nonlocal phonology of the language. We show how our model fits our experimental results as well as a broader range of restrictions in the literature on Aymara.

4.1 A model of learning projections from baseline phonology

4.1.1 A brief description of the learner

In this section, we present a brief description of our learning model, which is described in more detail in Gouskova and Gallagher (2017). The implementation of the model is available on GitHub at .

The model builds on Hayes and Wilson's (2008) UCLA Phonotactic Learner (UCLAPL). The first stage of the learning procedure constructs a phonotactic grammar based on a list of phonological words and features describing each segment (see (8)). The model proceeds to construct a set of constraints against unattested and underattested sequences (see (9)). These constraints are formulated in terms of natural classes, and they are given weights using a Maximum Entropy procedure, which seeks to maximize the probability of the learning data. The resulting grammar can be used to assign harmony scores to test words, so that its fit to the data can be compared against, e.g., experimental data from human speakers (see Goldwater and Johnson, 2003, Daland et al., 2011, Berent et al., 2012, Hayes and White, 2013, and Wilson and Gallagher 2018).

4.1.2 Building projections from cues in the baseline grammar

The original version of the learner has the capability to posit constraints on autosegmental projections provided by the analyst: for example, Hayes and Wilson (2008) demonstrate that their learner can find constraints enforcing vowel harmony in Shona verbs when it is given a projection that includes only vowels; the learner can also capture the stress pattern of Wargamay when given the appropriate projections for segments that bear primary and secondary stress. We exploit this capability in our extension to the learner, which finds projections and/or modifies the training data automatically when certain cues are present in the baseline grammar. For the simulations reported in this paper, two such cues are instrumental:

- (6) **Segmental placeholder trigrams:** constraints of the form *X-any_segment-Y, where X and Y are part of a natural class Z. When the learner finds such trigrams, it adds projection Z to its search space of constraints.
- (7) **Morpheme-boundary trigrams:** constraints of the form *X-non_morpheme_boundary-Y, where X and Y are part of a natural class Z. When the learner finds such trigrams, it adds a projection Z to its search space of constraints.

Hayes and Wilson’s learner automatically adds the feature [\pm word boundary] to every feature set in order to capture word edge phonotactics. The non-boundary segments of each language are then automatically part of the largest natural class, [-word boundary]. Since Hayes and Wilson’s learner has a bias toward broad natural classes, the learner will identify the constraints that refer to this class relatively early compared to other trigrams, provided the language offers support for them. Intuitively, the presence of constraints whose middle segment can be any of the segments in the language, *X-any_segment-Y, is a cue to the learner that the segments to either side of “any segment” interact nonlocally. The logic for non-morpheme-boundary trigrams is similar. If the grammar includes a constraint *X-non_morpheme_boundary-Y, this tells us that X and Y are permitted across a morpheme boundary but not across any intervening segment, so *X-any_segment-Y holds tautomorphemically.

Our extension of the learner uses constraints of this type to posit a projection that includes whichever natural class is the smallest class including both X and Y. For example, if the learner finds that [l] and [r] cannot occur across any segment (as in an idealized version of Latin, Steriade 1987, Cser 2010), it will posit a projection of liquids. More specifically, in the simulation of Quechua reported in Gouskova and Gallagher (2017), the learner’s baseline includes the constraint *[-cont,-son][-wb][+cg], or *STOP-ANY_SEG-EJECTIVE. The smallest class that includes all stops and ejectives is the natural class of stops, so the stop projection is added to the grammar and searched for constraints. This is schematically illustrated below.

(8) Input to the learner:

a. training data {<#pata#>, <#p'ata#>, <#p^hata#>, <#t'ampa#>, <#map'a#>, <#lama#>, ...}

b. feature set

	word_boundary	sonorant	continuant	cg	sg	syllabic
p	-	-	-	-	-	-
t	-	-	-	-	-	-
p'	-	-	-	+	-	-
p ^h	-	-	-	-	+	-
m	-	+	-	0	0	-
l	-	+	+	0	0	-
a	-	+	0	0	0	+
<#	+	0	0	0	0	0
#>	+	0	0	0	0	0

(9) Stage 1 output of the learner: baseline grammar

Constraint	Proj.	Weight	Comment	Violated by
*[+son,+cont]	default	0.4	*[l], [l] is rare. Low weight, violable	lama
*[-son][+son]	default	12	bigram constraint on consonant clusters; e.g., [pm] not seen in data	*apla
*[-son,-cont][-wb][+cg]	default	15	trigram stop-any_segment-ejective; e.g., [pak'a] not seen in data	*pat'a (but not *tamp'a)
...				

(10) Learner posits a projection for the smallest natural class that includes [-son,-cont] and [+cg]:

Projection	What is visible
baseline/default	t a m p' a
[-son,-cont]	t p'

(11) Stage 2 output of the learner: projection grammar

Constraint	Proj.	Weight	Comment	Violated by
*[+son,+cont]	default	0.4	*[l], [l] is rare. Low weight, violable	lama
*[-son][+son]	default	12	bigram constraint on consonant clusters; e.g., [pm] not seen in data	*apla
*[-son,-cont][+cg]	-son-cont	20	projection bigram constraint on plain-ejective seq's; e.g., [p...t'] not seen in data	*pat'a, *tamp'a
...				

Occasionally, the learner identifies more than one placeholder constraint, in which case we allow it to search each of the resulting projections for constraints if the natural classes they entail are distinct.

If the learner is trained on morphologically parsed data, it may detect constraints of the form *X-non_morpheme_boundary-Y. In our feature sets, all segments are [-morpheme boundary], but word edges and morpheme boundaries are [+morpheme boundary]. If the learner posits a constraint whose middle gram is [-morpheme boundary], this means that segmental trigrams of the form *X-any_segment-Y are underattested, but the trigram X-morpheme_boundary-Y is attested often enough to exclude it from the formulation of the constraint. This situation will arise in a language like Aymara that has few STOP-ANY_SEGMENT-EJECTIVE trigrams but a fair number of STOP-MORPHEME_BOUNDARY-EJECTIVE trigrams. The phonotactic restriction is cancelled in heteromorphemic contexts—the occurrence of an ejective in close proximity to a stop is a boundary signal (*Grenzsignal*) in the sense of Trubetzkoy (1939). Constraints of this type also cue our learner to construct a projection. The only difference is that when the learner is looking at morphologically parsed words, the morpheme boundary symbol will also be present on all projections.⁴ If morpheme boundaries can be present on a projection, they will separate the segments that would otherwise form a bigram, as shown in 12. Henceforth, for compactness, we will write morpheme boundaries as “+”, and the non-morpheme boundary class will be [-mb].

⁴Morpheme boundary symbols are the simplest implementation for this, but there are of course alternatives. For a discussion of theoretical and learnability issues, see Pyle 1972, McCarthy 1989, Beckman 1997, Adriaans and Kager 2010, Becker and Allen submitted, Kastner and Adriaans 2018 and others.

(12) Projections with morpheme boundaries

Projection	What is visible	
baseline/default	t a m p' a	p a n + t' a
[-son,-cont]	t p'	p + t'

4.2 Parameters manipulated by the analyst

There are several parameters that affect the learner’s ability to find generalizations. First, the segmental features determine whether the learner can group the segments into the right natural classes. The learner is sensitive to the size of the classes, as well as their overall number (see Hayes and Wilson 2008, Gouskova and Gallagher 2017). Since we were primarily interested in laryngeal restrictions, we selected a feature set that uses mostly privative features (specifically, [plain], [cg] and [sg]). Hayes and Wilson’s learner favors constraints whose natural classes mention as few features as possible, so privative features allow certain classes (e.g., plain stops) to be picked out more easily. See §4.3 for the full list.

The other parameters have an effect on the number of constraints induced, the length of segmental strings they scope over, and how closely the learner fits the grammar to the data. We were generous in the number of constraints we allowed the learner to discover, since this learner stops when it cannot identify any constraints that pass the selection criterion. The length of constraint strings on the segmental projection ranged from 1 (as in *[+cg]) to 3 (as in *[-syllabic][-syllabic][-syllabic], “no CCC clusters”); the length of constraint strings on higher projections ranged from 2 to 3. In addition to these two fairly simple parameters, there are several parameters that affect the fit of the model in various ways.

The first parameter is **gain** (Della Pietra et al., 1997, Wilson and Gallagher, 2018, Gouskova and Gallagher, 2017). The gain of a constraint C is proportional to the reduction in the Kullback-Leibler divergence between the current grammar and the grammar with C added, and the weights of all the other constraints unchanged. A constraint can only be added if its gain exceeds the threshold; the higher the gain, the harder it is to add new constraints. We have found more-over that gain can be lower when the training data sets are small and each datum is relatively informative, but larger data sets yield more sensible grammars when gain is higher. The second parameter we manipulated is **gamma**. This parameter scales the harmony score relative to the negative log probability in the objective function of the learner. Increasing gamma makes it less likely that constraints with very low weights will be learned; individual violations of constraints end up being more important. There are additional parameters that can be manipulated, such as the **Laplace regularizer** λ , whose function is to penalize constraints with large weights (Wilson and Gallagher 2018:615). We set λ to a small constant 0.00001.

4.3 Learning data and features

The corpus we used was based on the Aymara wordlist on the An Crúbadán project website (<http://crubadan.org>), described in §2. We created two versions of the corpus: an unsegmented list of phonological words (transcribed on the basis of the transparent orthography), and a segmented list with morpheme boundaries. Recall that we only used those words in the An Crúbadán list that contained the roots that also occur in the De Lucca (1987) dictionary. Since morpheme boundaries add to the overall length of each string, we had to filter the segmented word list to exclude words above a certain length.⁵ This left us with 46,164 words each in the segmented and unsegmented lists.

The feature set we used for all simulations is shown in Table 11. In addition to the phonemes of Aymara, the feature set contains the morpheme boundary “+”, the word boundary, and a special “copy” segment X, which has just one privative feature, [+copy]. As explained in Gouskova and Gallagher (2017), this copy notation is necessary because the learner does not implement algebraic notation in its constraint language (Berent et al. 2012). Thus, to allow the learner to distinguish between the allowed identical pairs of ejectives and the disallowed non-identical ones (recall §2.1), we transcribe words such as [tʰantʰa] as [tʰanXa].

⁵The Java learner has a technical limitation: it cannot handle strings longer than about 40 characters. The maximum string length is used in generating the “sample salad” of phoneme strings that the learner compares to the learning data in figuring out what is missing from the learning data, and it must do so in finite time, so the shorter the words in the learning data, the better. See Daland (2015) for more.

	long	syll	son	cont	cg	sg	plain	lab	dent	pal	vel	uv	rhot	lat	nas	lo	bk	hi
p	0	-	-	-	0	0	+	+	0	0	0	0	0	0	0	0	0	0
t	0	-	-	-	0	0	+	0	+	0	0	0	0	0	0	0	0	0
tʃ	0	-	-	-	0	0	+	0	0	+	0	0	0	0	0	0	0	0
k	0	-	-	-	0	0	+	0	0	0	+	0	0	0	0	0	0	0
q	0	-	-	-	0	0	+	0	0	0	0	+	0	0	0	0	0	0
p'	0	-	-	-	+	0	0	+	0	0	0	0	0	0	0	0	0	0
t'	0	-	-	-	+	0	0	0	+	0	0	0	0	0	0	0	0	0
tʃ'	0	-	-	-	+	0	0	0	0	+	0	0	0	0	0	0	0	0
k'	0	-	-	-	+	0	0	0	0	0	+	0	0	0	0	0	0	0
q'	0	-	-	-	+	0	0	0	0	0	0	+	0	0	0	0	0	0
p ^h	0	-	-	-	0	+	0	+	0	0	0	0	0	0	0	0	0	0
t ^h	0	-	-	-	0	+	0	0	+	0	0	0	0	0	0	0	0	0
tʃ ^h	0	-	-	-	0	+	0	0	0	+	0	0	0	0	0	0	0	0
k ^h	0	-	-	-	0	+	0	0	0	0	+	0	0	0	0	0	0	0
q ^h	0	-	-	-	0	+	0	0	0	0	0	+	0	0	0	0	0	0
s	0	-	-	+	0	0	0	0	+	0	0	0	0	0	0	0	0	0
ʃ	0	-	-	+	0	0	0	0	0	+	0	0	0	0	0	0	0	0
χ	0	-	-	+	0	0	0	0	0	0	0	+	0	0	0	0	0	0
h	0	-	-	+	0	+	0	0	0	0	0	0	0	0	0	0	0	0
m	0	-	+	-	0	0	0	+	0	0	0	0	0	-	+	0	0	0
n	0	-	+	-	0	0	0	0	+	0	0	0	0	-	+	0	0	0
ɲ	0	-	+	-	0	0	0	0	0	+	0	0	0	-	+	0	0	0
r	0	-	+	-	0	0	0	0	+	0	0	0	+	-	-	0	0	0
l	0	-	+	+	0	0	0	0	+	0	0	0	-	+	-	0	0	0
ʎ	0	-	+	+	0	0	0	0	0	+	0	0	-	+	-	0	0	0
j	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	-	-	+
w	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	-	+	+
i	-	+	+	+	0	0	0	0	0	0	0	0	0	0	0	-	-	+
u	-	+	+	+	0	0	0	0	0	0	0	0	0	0	0	-	+	+
a	-	+	+	+	0	0	0	0	0	0	0	0	0	0	0	+	-	-
e	-	+	+	+	0	0	0	0	0	0	0	0	0	0	0	-	-	-
o	-	+	+	+	0	0	0	0	0	0	0	0	0	0	0	-	+	-
i:	+	+	+	+	0	0	0	0	0	0	0	0	0	0	0	-	-	+
u:	+	+	+	+	0	0	0	0	0	0	0	0	0	0	0	-	+	+
a:	+	+	+	+	0	0	0	0	0	0	0	0	0	0	0	+	-	-
e:	+	+	+	+	0	0	0	0	0	0	0	0	0	0	0	-	-	-
o:	+	+	+	+	0	0	0	0	0	0	0	0	0	0	0	-	+	-

Table 11: Features for Aymara computational simulations

4.4 The simulations

4.4.1 Baseline grammar trained on a corpus of segmented words

We trained the learner on the corpus of morphologically segmented words, since we expected it to be able to identify the generalizations about tautomorphic stops when it was supplied with the crucial information about morpheme boundaries. The simulation we report here had a gain of 400 and gamma of 150, and the grammar included 120 constraints. The baseline grammar contains three morphological trigram constraints (see Table 12), corresponding to the three restricted laryngeal combinations. Plain-ejective, plain-aspirate and ejective-ejective combinations are found across a morpheme boundary but are underattested across any intervening segment.

	*X-non_morpheme_boundary-Y	Weight	Sequences penalized
a.	*[+plain][-mb][+cg]	13.214	[p,t,k,tʃ,q]-seg-[p',t',k',tʃ',q']
b.	*[+plain][-mb][-cont,+sg]	13.039	[p,t,k,tʃ,q]-seg-[p ^h ,t ^h ,k ^h ,tʃ ^h ,q ^h]
c.	*[+cg][-mb][+cg]	12.886	[p',t',k',tʃ',q']-seg-[p',t',k',tʃ',q']

Table 12: Placeholder constraints in the baseline grammar trained on segmented words

These three constraints act as cues to the creation of two projections, shown in (13). The oral stop projection is motivated by constraints (a) and (b), since the smallest natural class grouping over plain stops and aspirates or over plain stops and ejectives is the class of all oral stops. The ejective projection is motivated by constraint (c).

- (13) Projections posited from examining the new baseline grammar built from the monomorphemic sublexicon

Projection	defining features	What is visible
oral stops	[-son,-cont]	p t k tʃ q, p' t' k' tʃ' q', p ^h t ^h k ^h tʃ ^h q ^h , +
ejectives	[+cg]	p' t' k' tʃ' q', +

4.4.2 The full grammar with projections

In the next step of the learning procedure, the same training data set is revisited with the two projections identified in the baseline simulation, in addition to the default projection.

Table 13 shows all of the constraints that the learner posited on the two nonlocal projections. The constraints are shown in the order they were added to the grammar—following Hayes and Wilson’s heuristics, bigram constraints are considered first, since there are fewer of them than trigrams. There are several constraints on the classes of ejectives and aspirates, as well as constraints on individual ejective and aspirate segments. There are also some constraints on place

of articulation combinations. The constraints that capture the restrictions on plain-ejective and ejective-ejective combinations are given in bold.

Projection	Constraint	Weight	Sequences penalized
a. [+cg]	*[-syll][-syll]	4.899	ejective...ejective
b. [+cg]	*[-wb][+palatal][+wb]	11.938	ejective/+...[tʃ]...#
c. [+cg]	*[-wb][+uvular][+wb]	11.324	ejective/+...[q']...#
d. [-son, -cont]	*[-wb][+cg,+labial]	12.694	stop/+...[p']
e. [-son, -cont]	*[-syll][+cg,+velar]	11.629	stop...[k']
f. [-son, -cont]	*[-syll][+cg,+uvular]	11.274	stop...[q']
g. [-son, -cont]	*[+dental][+palatal]	12.616	[t t' t ^h]...[tʃ tʃ' tʃ ^h]
h. [-son, -cont]	*[+velar][+uvular]	11.673	[k k' k ^h]...[q q' q ^h]
i. [-son, -cont]	*[+plain,+labial][+sg]	12.352	[p]...aspirate
j. [-son, -cont]	*[+plain,+uvular][+sg]	11.298	[q]...aspirate
k. [-son, -cont]	*[+plain,+dental][+sg]	11.768	[t]...aspirate
l. [-son, -cont]	*[+uvular][+velar]	5.629	[q q' q ^h]...[k k' k ^h]
m. [-son, -cont]	*[-wb,+mb][+sg,+palatal]	12.242	+...[tʃ ^h]
n. [-son, -cont]	*[-wb][-syll][+sg]	11.366	stop/+...stop...aspirate
o. [-son, -cont]	*[+plain][+cg][+wb]	6.351	plain stop...ejective...#
p. [-son, -cont]	*[-syll][-syll][+uvular]	11.594	stop...stop...[q q' q ^h]
q. [-son, -cont]	*[-syll][-syll][+palatal]	11.628	stop...stop...[tʃ tʃ' tʃ ^h]
r. [-son, -cont]	*[+cg][-syll][-syll]	11.832	ejective...stop...stop

Table 13: Part of the final grammar induced from training on the full segmented corpus of Aymara: Constraints discovered on the nonlocal projections

To assess how these constraints capture the phonological restrictions in Aymara, we test how the grammar rates the nonce words from the repetition studies in §3 and then go on to look at a broader set of structures.

4.4.3 Testing the model against experimental results

The grammar makes many of the same distinctions among nonce words that Aymara speakers made in the repetition experiment. The correlations between repetition accuracy and the score assigned by the grammar are plotted in Figures 3 and 4.

Figure 3 pools accuracy averages across participants for each nonce word in Experiment 1; the regression line shows the overall correlation between repetition accuracy and harmony scores (the shaded region is the 95% confidence interval). Each data point in the plot represents an average accuracy score for a specific word in the experiment, labeled according to the type of stimulus: “CT” is control, “PE” is plain-ejective, “EE” is ejective-ejective. Like participants in the experiment, the model distinguishes control words from forms with the restricted plain-ejective

and ejective-ejective combinations. The model assigns a score of -6 to all control words, compared to an average score of -22 to plain-ejective or ejective-ejective forms.

The model also reflects the distinction between dental ejectives and other ejectives. The grammar includes constraints that penalize all ejective-ejective or plain-ejective combinations, but it also includes specific constraints on stop. . . [p’], stop. . . [q’], and stop. . . [k’] combinations, which further penalize forms with non-dental ejectives. Plain-ejective and ejective-ejective combinations with a dental ejective receive a higher average score of -12, while these same laryngeal combinations with other medial ejectives receive a lower average score of -25.

In the experiment, participants’ accuracy was somewhat higher for ejective-ejective forms (50%) than plain-ejective forms (32%)—a distinction that is not reflected in the model (both categories have an average score of -22). This difference was small, though significant, in the behavioral data, but the difference between plain-ejective and ejective-ejective forms was inconsistent across the two experiments, so we cannot draw firm conclusions about differences in grammaticality between these two types of combinations. The correlation between the model’s harmony scores and the Aymara speakers’ average accuracy in repeating the words is fairly high (Kendall’s $\tau=0.68$, Spearman’s $\rho=0.83$). We report non-parametric correlations only because the harmony scores are not normally distributed.

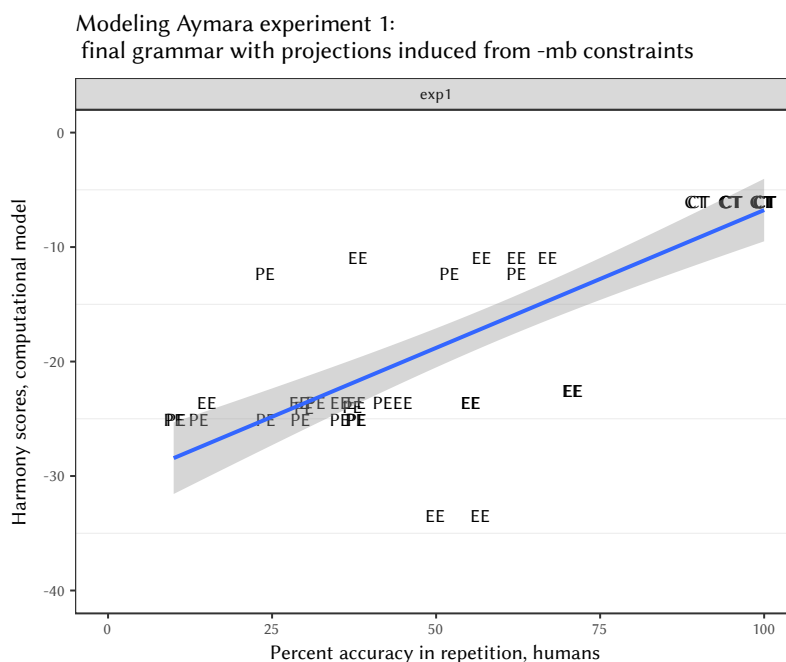


Figure 3: Harmony scores for stimuli from repetition experiment 1, assigned by the final grammar trained on segmented word corpus. Each point in the plot is labeled according to stimulus type: “CT” is control, “PE” is plain-ejective, “EE” is ejective-ejective.

Fig. 4 plots the nonce words tested in Experiment 2. The stimulus types are labeled as follows: ED: ejective-ejective-dental, EL: ejective-ejective-labial, PD: ejective-plain-dental, PL: ejective-plain-labial. The model assigns higher scores to forms with a dental ejective (-12 on average for both ejective-dental and plain-dental combinations), which are polymorphemic, than to monomorphemic forms with a labial ejective (-33 for both ejective-labial and -28 plain-labial combinations), reflecting the main effect of place of articulation in the experiment (overall accuracy on forms with dental ejectives: 74%; overall accuracy on forms with labial ejectives: 35%). Again, the small, inconsistent differences between ejective-ejective and plain-ejective combinations observed in the experiments are not reflected in the model. The correlations between harmony scores assigned by the model and the Aymara speakers' average accuracy in this experiment are $\tau=0.52$, $\rho=0.72$.

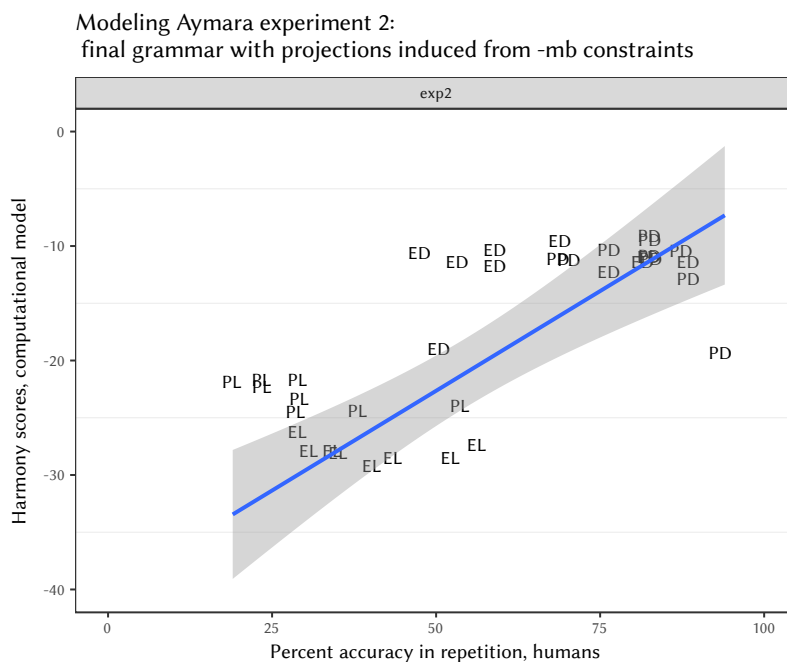


Figure 4: Harmony scores for stimuli from repetition experiment 2, assigned by the final grammar trained on the segmented word corpus. Each point is labeled according to stimulus type: ejective-ejective (ED: ejective-dental, EL: ejective-labial), plain-ejective (PD: plain-dental, PL: plain-labial).

The model captures the distinction between labial and dental forms in two ways. First, forms with labials have a tautomorphemic plain-ejective or ejective-ejective combination, so they violate the constraints $*[+plain][+cg][+wb]$ on the stop projection or $*[-syllabic][-syllabic]$ on the ejective projection. Forms with dental ejectives, on the other hand, have a morpheme boundary intervening between the stops and thus escape a violation of these constraints. Second, the model contains a constraint on labial ejectives that are not the first stop in the word ($*[-wb][+labial,$

+cg]), but no such constraint on dental ejectives, resulting in lower scores for forms with a labial ejective than a dental ejective, regardless of morphological structure.

As discussed in §3.3, the independent roles of place of articulation and morphological structure can be teased apart by comparing the results of Experiment 1 and Experiment 2. In the behavioral data, there were fewer errors on forms with dental ejectives in both experiments, but the difference between forms with dental and non-dental ejectives was larger in Experiment 2, where morphological structure was also at play. Our model shows the same qualitative pattern, though the difference is slight: the difference in scores between dental and non-dental forms is 13 points in Experiment 1 but 18 points in Experiment 2.

4.4.4 Beyond the experimental data: evaluating the full range of stop restrictions

We have focused thus far on just two underattested structures in Aymara: plain-ejective and ejective-ejective combinations. There are other underattested stop combinations in Aymara, however, and in this section we look at how our model reflects these other restrictions.

As discussed in §2, plain-aspirate sequences are underattested. Our model includes four constraints that penalizes plain-aspirate combinations. There are constraints against three specific plain stops followed by the class of aspirates - [p]-aspirate, [t]-aspirate and [q]-aspirate - as well as the more general constraint [-wb][-syllabic][+sg]. This latter constraint penalizes all stop-aspirate combinations, not just plain-aspirate ones, but only when they are preceded by another stop or a morpheme boundary.

To assess the grammar, we constructed a small set of targeted test words. All test words had a CaCa structure, and contained two stops. Table 14 shows the scores that our model assigns to words with several different laryngeal configurations: four (a)–(d) that are described as unrestricted in the literature (ejective-aspirate and aspirate-ejective combinations are discussed below), and three that are restricted (e)–(g). Scores in this and subsequent tables were averaged only over words that did not violate any of the other restrictions described in the rest of this section, to allow assessment of each restriction individually. The table shows that the grammar clearly distinguishes between unrestricted combinations, which receive a high score of -6, and the restricted combinations, which receive lower scores. The plain-ejective and ejective-ejective combinations receive lower scores than plain-aspirate combinations because not all plain-aspirate combinations are penalized by the grammar. The model only penalizes forms with three of the plain stops, [p t q], followed by aspirates, but assigns a score of -6, comparable to unrestricted combinations, to forms with [ʃ]-aspirate or [k]-aspirate combinations. In this case, the model dances around the exceptions to the restriction, positing several more specific but more accurate constraints on individual plain-aspirate combinations as opposed to a single more general but less accurate constraint covering all plain-aspirate combinations.

Laryngeal combination	Description in literature	Harmony score
a. pl-pl	unrestricted	-6
b. ej-pl	unrestricted	-6
c. asp-pl	unrestricted	-6
d. asp-asp	unrestricted	-8
e. pl-ej	restricted	-20
f. pl-asp	restricted	-16
g. ej-ej	restricted	-23

Table 14: Harmony scores assigned by our final model to a small test set of nonce words assessing laryngeal restrictions

Aymara shows place co-occurrence restrictions as well as laryngeal co-occurrence restrictions. Combinations of dorsals (uvulars and velars) and combinations of coronals (dentals and palatals) are infrequent. Our model includes the constraints *[dental][palatal], *[velar][uvular] and *[uvular][velar] on the stop projection, which penalize three of the four possible combinations. The distinction between dental-palatal and palatal-dental combinations seems reasonable: there are 84 palatal-dental sequences in our corpus compared to 24 dental-palatal sequences. For dorsals, there is just 1 uvular-velar combination and 0 velar-uvular combinations in our corpus, so both constraints are warranted. Table 15 shows that dental-palatal, velar-uvular and uvular-velar combinations receive low scores, while palatal-dental combinations receive higher scores comparable to the unrestricted laryngeal combinations.

Type of restriction	score
a. dental-palatal	-24
b. palatal-dental	-6
c. velar-uvular	-18
d. uvular-velar	-14

Table 15: Harmony scores assigned by our final model to a small test set of nonce words assessing place co-occurrence restrictions

A set of quite complicated restrictions applies to ejective-aspirate and aspirate-ejective pairs. While both of these laryngeal combinations are attested, not all individual combinations of segments occur. The attested combinations of ejectives and aspirates in our corpus are shown in Table 16. As described in MacEachern (1997), which stop is ejective and which is aspirate is predictable based on place of articulation. If the initial consonant is a labial or a uvular, it will be aspirated (see (j)–(l)); otherwise, it is ejective (see (a)–(h)). In uvular-labial pairs, the uvular is ejective (see (i)). Any combination not shown in the table is unattested.

ejective-aspirate	observed	ejective-aspirate	observed	aspirate-ejective	observed
a. t'...p ^h	2	f. tʃ'...q ^h	5	j. p ^h ...t'	7
b. t'...k ^h	11	g. k'...p ^h	15	k. p ^h ...tʃ'	24
c. t'...q ^h	334	h. k'...t ^h	11	l. q ^h ...t'	7
d. tʃ'...p ^h	29	i. q'...p ^h	13		
e. tʃ'...k ^h	56				
<i>Total</i>			476	<i>Total</i>	38

Table 16: Ejective and aspirate counts

Table 17 shows the harmony scores assigned to nonce words with attested and unattested combinations of ejectives and aspirates. The model correctly distinguishes between attested and unattested aspirate-ejective sequences, via the three place specific constraints *[-wb][+labial, +cg], *[-wb][+uvular, +cg] and *[-wb][+velar, +cg] which penalize non-initial labial, velar and uvular ejectives that are preceded by another stop. The grammar doesn't include any constraints on ejective-aspirate sequences, and attested and unattested combinations are only weakly distinguished by the model. This difference arises from an orthogonal bigram constraint against [aeo][tʃ^h] sequences.

	Combination	Harmony score
a.	aspirate-ejective, attested	-6
b.	aspirate-ejective, unattested	-16
c.	ejective-aspirate, attested	-7
d.	ejective-aspirate, unattested	-10

Table 17: Harmony scores assigned by our final model to nonce words for aspirate/ejective combinations broken down by place

Hayes and Wilson's model has a preference for more general constraints, stated over larger natural classes. To completely match the distribution of stops in the language, the model would have to include many constraints on individual segmental combinations. While the grammar does include several constraints that refer to classes of a single segment, in other cases such specific constraints are not learned. We leave it to future experimental work to identify what generalizations Aymara speakers have learned about ejective-aspirate and aspirate-ejective combinations, and how the model might need to be modified to match speaker behavior.

Finally, pairs of segments that differ only in laryngeal features (e.g., [t'...t^h], [p'...p], [k^h...k], etc.) are also reported to be restricted and are nearly absent in our corpus. Our model does not include constraints on any of these combinations, again because such constraints would refer to individual segments and the learner typically does not learn such constraints.

4.4.5 An unsegmented corpus

To further establish the role of morphological information in the success of our model, we ran learning simulations on the same word corpus, but with morpheme boundary markers removed. Recall from Table 3 that there are many instances of plain-aspirate, plain-ejective and ejective-ejective combinations in the corpus as a whole, but they are mostly found across a morpheme boundary. We tested whether these sequences were frequent enough to obscure the restrictions if morpheme boundaries are not represented.

We first examined whether our model found placeholder trigram constraints when trained on the unparsed corpus. We tested a range of gain and gamma combinations, and compared the placeholder trigrams found on the baseline model when trained on the parsed and unparsed data sets. All models were asked to find a maximum of 200 constraints. We report on a representative sample of the numerous combinations we tried in Table 18. Morpheme boundary trigram constraints corresponding to at least two of the three restricted combinations are found for the parsed data under almost all settings, and settings with a higher gain or gamma allow the model to detect all three restrictions in the baseline grammar. For the unparsed data, placeholder trigram constraints are only found with a very low gamma of 1. Even with gamma this low, the model only finds a single placeholder trigram on ejectives; the model never finds a placeholder trigram corresponding to the plain-aspirate restriction and may miss the ejective-ejective restriction as well.

gain	gamma	parsed corpus	unparsed corpus
100	1	<i>none</i>	*[+plain][][+cg]
400	1	*[+plain][-mb][-continuant,+sg]	*[+plain][][+cg]
		*[-sonorant,-continuant][-mb][+cg]	
800	1	*[+plain][-mb][-continuant,+sg]	*[-son, -cont][][+cg]
		*[-sonorant,-continuant][-mb][+cg]	
100	50	*[+plain][-mb][+cg]	<i>none</i>
		*[+plain][-mb][+sg, -cont]	
400	50	*[+plain][-mb][+cg]	<i>none</i>
		*[+plain][-mb][+sg, -cont]	
800	50	*[+cg][-mb][+cg]	<i>none</i>
		*[+plain][-mb][+cg]	
		*[+plain][-mb][+sg, -cont]	
100	150	*[+cg][-mb][+cg]	<i>none</i>
		*[+plain][-mb][+cg]	
		*[+plain][-mb][+sg, -cont]	
400	150	*[+cg][-mb][+cg]	<i>none</i>
		*[+plain][-mb][+cg]	
		*[+plain][-mb][+sg, -cont]	
800	150	*[+cg][-mb][+cg]	<i>none</i>
		*[+plain][-mb][+cg]	
		*[+plain][-mb][+sg, -cont]	

Table 18: Placeholder trigrams found at various settings in the parsed and the unparsed versions of the Aymara corpus

One of the grammars built after training on the unparsed corpus is investigated in more detail in Fig. 5 and Table 19. This grammar’s gain is 500, and gamma=1. The baseline grammar constraint *[+plain]-any_seg-[+cg] motivates the [-son, -cont] projection, on which all restrictions could be correctly stated. But the final grammar achieves a poor fit to the experimental data, as shown by the nearly flat lines in Fig. 5. The correlations for Experiment 1 are $\tau=0.22$ and $\rho=0.35$; the correlations for Experiment 2 are $\tau=0.43$ and $\rho=0.60$.

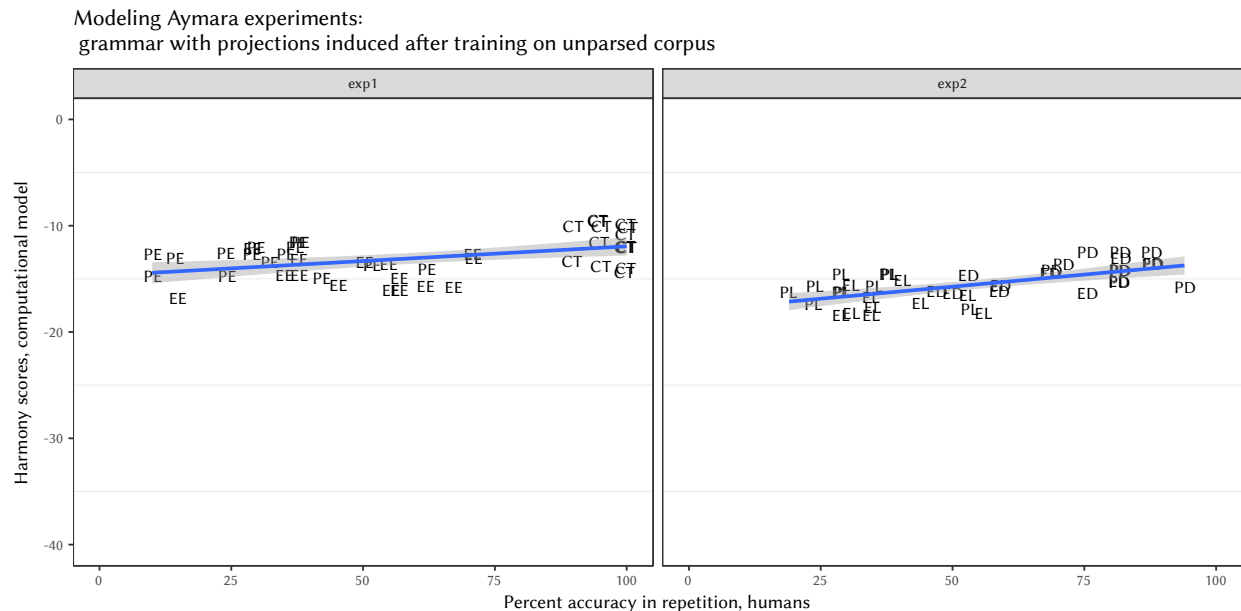


Figure 5: A grammar with induced projections built from the unsegmented corpus, tested on Aymara experimental data. Data points are labeled according to stimulus type; Exp. 1: “CT” is control, “PE” is plain-ejective, “EE” is ejective-ejective. Exp. 2: co-occurrence (ED: ejective-dental, EL: ejective-labial), ordering (PD: plain-dental, PL: plain-labial)

The model’s lack of success at representing phonologically meaningful underattestations in the data becomes clear when we look at the constraints it posits on the stop projection (see Table 19). Their low weights are due in part to the gamma setting; almost all of them are violated frequently by Aymara words. While these constraints penalize some restricted combinations, they don’t capture the full extent of the restrictions nor are their weights high enough to distinguish restricted from unrestricted structures.

Constraint on -son,-cont projection	weight	violated by
a. *[-wb][+cg]	1.494	stop...[p' t' k' q' tʃ']
b. *[+cg][+wb]	1.998	[p' t' k' q' tʃ'] ... #
c. *[+plain,+uvular][+wb]	0.989	[q q ^h q'] ... #
d. *[+palatal][+wb]	1.154	[tʃ tʃ ^h tʃ'] ... #
e. *[+cg,+labial][]	2.068	[p p ^h p'] ... stop
f. *[+velar][+uvular]	1.73	[k k ^h k'] ... [q q ^h q']
g. *[+plain][+sg,+uvular]	4.31	[p t k q tʃ] ... [q ^h]
h. *[+cg,+uvular][-wb]	0.551	[q'] ... stop
i. *[+wb][+cg,+dental][+palatal]	2.731	# [t'] [tʃ tʃ ^h tʃ']
j. *[+cg,+palatal][-wb]	0.997	[tʃ'] ... stop
k. *[+cg,+velar][][+wb]	1.981	[k'] ... stop ... #

Table 19: Constraints on the stop projection discovered after training on a corpus of unparsed words

With this low gamma setting, the model does not succeed in distinguishing meaningful underattestations in the data, and instead learns many, low weighted constraints with numerous exceptions. Our more succesful model in §4.4.3 has higher gamma and gain, which make the model more selective and a better fit to the phonological distinctions supported by traditional phonological analysis and experimental work with native speakers. Given these same settings (400 gain, 150 gamma), a grammar trained on the unparsed corpus fails to include any placeholder trigram constraints from which it could posit nonlocal projections.

We also considered whether it was possible to capture phonological distinctions on the stop projection with unparsed data, when the model was given a higher gain and gamma and we supplied the stop projection manually. As shown in Figure 6, this grammar achieves a better fit to Aymara speakers' performance in the repetition experiments than the grammar in Fig. 5. But there are interesting differences in the details.

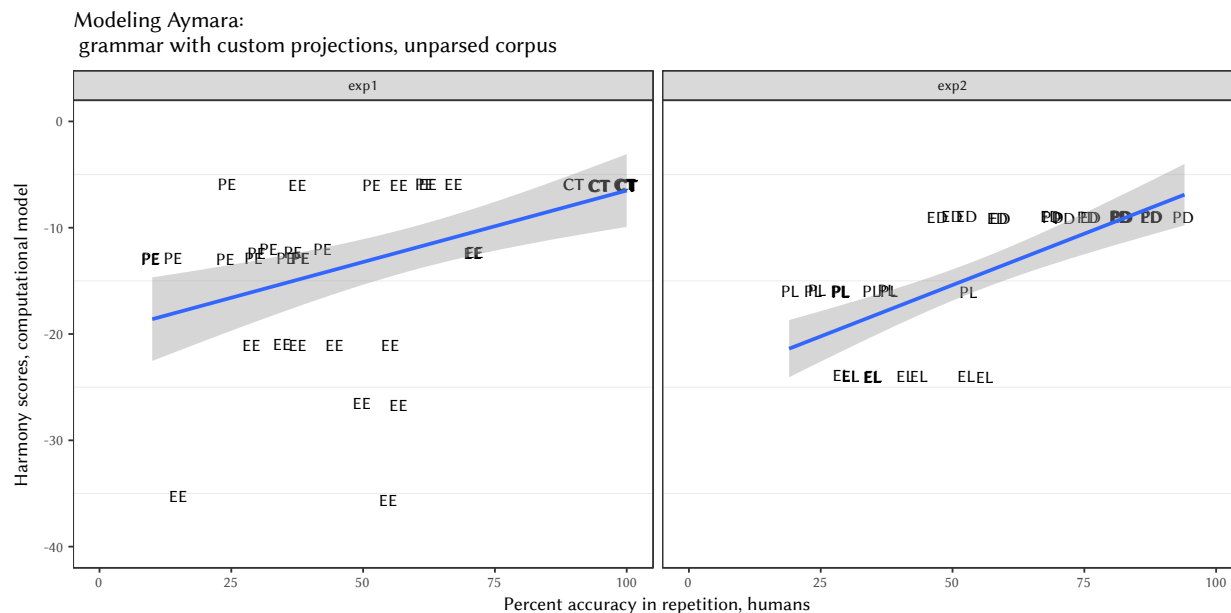


Figure 6: A grammar built from the unsegmented corpus, tested on Aymara experimental data, with manually supplied projections. Data points are labeled according to stimulus type; Exp. 1: “CT” is control, “PE” is plain-ejective, “EE” is ejective-ejective. Exp. 2: co-occurrence (ED: ejective-dental, EL: ejective-labial), ordering (PD: plain-dental, PL: plain-labial)

The unparsed model captures the distinction between dental and labial ejectives via constraints on everything but dentals (in Table 20, (c), (e)–(g), (k)–(m)). But this model fails to make a distinction between control stimuli such as [lap’a] (100% correct) and [p’it’a] (67% correct)—they all receive a harmony score of -6. This is because this grammar does not include a general constraint against ejectives in 2nd position on the stop projection. Constraint (c) on the ejective position is too specific. This grammar is overfitting, learning overly specific constraints to accommodate the exceptions in the data.

When it comes to Experiment 2, however, the model’s fit to behavioral data is comparable to the model we reported in §4.4.3 (although the differences between “good” and “bad” forms are smaller in Experiment 2—the opposite of the parsed grammar). The reasons for this have to do with the abundance of dental ejectives in Aymara suffixes; by positing place-specific constraints against non-dental ejectives in 2nd position on the stop projection, the model manages to approximate the same generalizations.

	Projection	Constraint	Weight	Seq's penalized
a.	+cg	*[-wb][+labial]	8.06	ejective...[p']
b.	+cg	*[-wb][+palatal]	5.531	ejective...[tʃ]
c.	+cg	*[+dental][-wb]	14.559	[t']...ejective
d.	-son-cont	*[][-wb,+mb]	17.915	stop...+
e.	-son-cont	*[-wb][+cg,+uvular]	6.023	stop...[q']
f.	-son-cont	*[-wb][+cg,+velar]	6.265	stop...[k']
g.	-son-cont	*[-wb][+cg,+labial]	6.942	stop...[p']
h.	-son-cont	*[+plain][+sg,+uvular]	4.974	[p t k q tʃ]...[qʰ]
i.	-son-cont	*[+dental][+sg,+palatal]	6.33	[t tʰ t']...[tʃʰ]
j.	-son-cont	*[+dental][+cg,+palatal]	13.976	[t tʰ t']...[tʃʰ]
k.	-son-cont	*[+plain,+labial][+cg,+palatal]	13.507	[p]...[tʃʰ]
l.	-son-cont	*[+palatal][+cg,+palatal]	5.633	[tʃ tʃʰ tʃʰ]...[tʃʰ]
m.	-son-cont	*[+plain,+uvular][+cg,+palatal]	12.905	[q]...[tʃʰ]
n.	-son-cont	*[+cg,+labial][+sg]	5.42	[p']...aspirate
o.	-son-cont	*[+plain][+sg][+cg]	13.224	plain...aspirate...ejective
p.	-son-cont	*[+plain][+sg][+sg]	13.625	plain...aspirate...aspirate
q.	-son-cont	*[-wb][] [+cg,+palatal]	13.283	stop...stop...[tʃʰ]
r.	-son-cont	*[+sg][+plain][+sg,+labial]	12.555	aspirate...plain...[pʰ]
s.	-son-cont	*[+sg][+cg][+sg]	12.723	aspirate...ejective...aspirate

Table 20: Constraints on ejective and stop projections induced from a corpus without morpheme boundaries

For a quantitative comparison, Table 21 summarizes the non-parametric correlations between the harmony scores each model assigns to the experimental stimuli and the averaged accuracy in the repetition experiments with Aymara speakers. Model (a), which uses parsed data both for inducing the projections and for the final grammar has the best correlations with Experiment 1, and the best correlations overall. Model (b), which includes no morphological information at all, achieves the lowest correlations across the board. The third model, (c), does worse on the first experiment and slightly better on the second experiment, but its overall correlations with behavioral data are not as good as the model in (a).

	Experiment 1		Experiment 2		Overall	
	τ	ρ	τ	ρ	τ	ρ
a. Induced Parsed Model (Figs. 3, 4)	0.68	0.83	0.52	0.72	0.59	0.77
b. Induced Unparsed Model (Fig. 5)	0.22	0.35	0.43	0.60	0.31	0.42
c. Manual Unparsed Model (Fig. 6)	0.49	0.67	0.55	0.74	0.52	0.69

Table 21: Correlations between harmony scores assigned by the three models and accuracy in repetition experiments with Aymara speakers

Zooming out from laryngeal co-occurrence restrictions, the model trained on the unparsed

corpus is not quite right in other ways. Aymara morphemes obey an exceptionless constraint against CCC clusters (0 of them in the corpus), but such clusters are created by syncope at morpheme boundaries. The right constraint to capture this would be *[-syll][-syll][-syll]. The Induced Parsed Model (a) contains such a constraint, and gives it a high weight of 14.506. The Manual Unparsed Model cannot motivate such a constraint—there are 3322 words in the Aymara corpus that have such clusters. What this model does instead is posit many specific constraints, sometimes with rather low weights, against various CCC clusters that it sees few examples of—e.g., *[-son,-cont][+labial][-cont], with a weight of 6.778. This is just one example of a morpheme structure constraint that a morphology-agnostic learner cannot capture.

5 Discussion

5.1 Morpheme boundaries and place of articulation

The experimental and modeling work show that the distinction between tautomorphemic and heteromorphemic stop-ejective sequences is likely both a direct effect of morphology and also an effect of place of articulation. Participants in Experiment 1 made slightly fewer errors on stimuli with dental ejectives than with other ejectives, reflecting the frequency of dental ejectives in non-initial position. This effect was exaggerated in Experiment 2, where the structure of nonce words favored a polymorphemic parse. Similarly, the constraints in the grammar are sensitive to both morphological structure and place of articulation. The model includes constraints on tautomorphemic but not heteromorphemic laryngeal combinations and it also includes more specific constraints on individual segments; due to their presence in suffixes, [tʰ] and [tʰʷ] are more frequent than other ejectives in non-initial position, and the constraints in the grammar reflect this asymmetry.

One element of the experimental results not directly captured by the model is that while participants were more accurate on dental ejectives than labial ejectives in Experiment 2, they still made more errors on dental ejective forms than on filler items. In contrast, our model predicts forms like [kʰastʰaŋa] to be fully grammatical. Errors on dental ejectives likely reflect the two available morphological parses for these forms. Since these were nonce words that weren't associated with any meaning, speakers did not know for certain that [tʰa] in these forms was the verbal suffix. There are also roots in the language that end in [tʰa]. Errors in the repetition task are predicted if participants occasionally parse a form like [kʰastʰaŋa] as [kʰastʰa+ŋa], while accurate repetitions are expected if the [kʰas+tʰa+ŋa] parse is hypothesized.

5.2 Phonotactic learning and segmentation

Our modeling simulations showed that while some pieces of the restrictions can be detected in an unsegmented corpus, morpheme boundaries are necessary to fully capture the patterns. This means that infants and children learning Aymara cannot have a complete phonotactic grammar until they have learned enough morphology to segment the speech stream. The prediction is that the trajectory of phonotactic awareness of laryngeal restrictions should be different in Aymara learners than in learners of a language like Quechua, where laryngeal restrictions are categorical and hold at the word level.

Even though languages do not mark every morpheme boundary phonotactically, speakers of languages such as Finnish and Dutch have been shown to use phonotactic knowledge to segment speech in experimental settings (Suomi et al., 1997, McQueen, 1998). The model we presented makes a simplifying assumption that at some point, the learner examines a fully parsed corpus with boundaries. A more realistic approach would use phonotactics to deduce where boundaries are located (as in the StaGe model of Adriaans and Kager 2010). StaGe uses bigram probabilities to posit word boundaries. Aymara would lend itself to such an approach: ejectives and aspirates are most common root-/word-initially (recall Table 2), so the distribution of plain-ejective bigrams would be a clue to boundaries even for a learner that does not yet have detailed morphological segmentation information. We leave for future work an implementation of a more complete model that deduces both where morpheme/word boundaries are and whether they lead to nonlocal projections.

5.3 Morphologically sensitive phonotactics and the subset problem

Patterns such as those of Aymara present two types of subset problem (Baker 1979, Bowerman 1988). First, phonotactic learning in general requires the learner to err on the side of assuming more restrictive grammars and to construct its own negative evidence. In order to posit these more restrictive grammars, an Optimality-Theoretic learner with innate constraints requires a bias to keep faithfulness constraints ranked low; the negative evidence comes from the theory's GEN component (Prince and Tesar 2004, Hayes 2004). If the learner induces constraints from data instead, it constructs its own negative evidence by comparing the attested data to plausible phonotactic distributions generated at random (Hayes and Wilson 2008). But even such a learner will not notice nonlocal interactions—it must either be given the nonlocal representations a priori (as in Hayes and Wilson's proposal), or it must be nudged in the direction of looking for nonlocal representations. In our proposal, the learner does a second pass of phonotactic learning in response to generalizations that it may have missed on the first pass, signaled by placeholder trigram constraints. This is a kind of bias is designed to alert the learner to the need for more

restrictive constraints.

Second, if the learner assumes that the sequences allowed in words are also allowed in morphemes, then this morphologically agnostic learner will learn the superset grammar. For a language such as English, the superset learner hears words with [md] clusters and assumes, incorrectly, that such clusters are allowed anywhere, not just at morpheme boundaries. Our experiments suggest that Aymara speakers make the more conservative assumption. We suggest that in order to learn the right level of generalization, an Aymara learner has to revisit phonotactic learning once morphological information is available. We implemented this by supplying morpheme boundaries and using sequences at boundaries as a clue that nonlocal interactions are present.

The alternative we did not discuss is to split the learning data into morphemes, and learn phonotactics over these morphemes. For Aymara, a plausible learning data set that would reveal the right regularities would be a corpus of roots. This is the learning data we used in Gouskova and Gallagher (2017). The benefit of using roots as learning data is that the learner may use just one simple cue, segmental placeholder trigrams, without attending to morpheme boundaries or reifying them to the representational level of segments with feature values. This would essentially use a *sublexicon* to learn morpheme-level phonotactics that hold over just a subset of the language's forms (Gouskova and Becker 2013, Becker and Gouskova 2016). The main reason we did not use a sublexicon model here is that it is not clear how to define phonotactics over bound roots; verbal roots in Aymara are obligatorily suffixed. The application of phonotactic learning to words is straightforward to implement, whether they are embedded in connected speech or taken as a lexicon-like list. On the other hand, bound roots and suffixes are not complete, pronounceable phonological objects, and learning over such entities would be one level of abstraction removed from a realistic learning scenario.

5.4 Cue-based learning

Our approach uses properties of the learning data to detect other properties of the language, and as such, it can be considered to be an example of cue-based learning (Dresher and Kaye 1990, Gibson and Wexler 1994, Dresher 1999). A critique of cue-based learning is that it assumes a lot of learning machinery specific to language (Nazarov and Jarosz 2017). We do believe that the problem of boundary-sensitive phonotactics is a fairly language-specific one, and it is not immediately clear how one would approach it without recognizing morphemes as separate pieces with boundaries.

The logic of inducing nonlocal interactions from trigrams is not strictly phonological: one could argue for a domain-general status of the deduction that if A and B cannot co-occur nearby in

a configuration A-X-B, it worthwhile to check whether A and B can co-occur at longer distances. Our learner uses an independent criterion to check whether A and B can interact at all—they must be part of a natural class, as defined by the language’s phonological contrasts and alternation system. This requirement, along with the requirement that constraints in the grammar must receive robust statistical support, minimizes the learner’s ability to notice nonlocal interactions between unrelated segments, which most linguists would describe as accidental.

6 Conclusion

This paper has examined a set of morphologically sensitive, nonlocal restrictions in Aymara in a corpus study, behavioral experiments, and a computational model. Our corpus results showed that the restricted combinations are not categorical, even tautomorphemically, though there is an asymmetry between tautomorphemic and heteromorphemic combinations. We show that both the nonlocality and the morphological sensitivity of the restrictions is observable from trigram constraints in a grammar trained on just the linear string of segments. By building projections based on these morpheme boundary trigram constraints, our model captures the range of restrictions reported in the language. The induced grammar succeeds through a combination of general constraints on relatively large classes, like the class of stops or the class of ejectives, and more specific constraints on individual segments.

Our experimental work supports the traditional description of the phonotactics of the language, and our modeling work shows that nonlocal restrictions are learnable inductively, by attuning to properties of the phonotactics of the linear string. By comparing a phonotactic grammar trained on parsed vs. unparsed data, we saw that the patterns are largely obscured by the exceptions found across morpheme boundaries—only one of the restricted combinations is observable as a baseline trigram in the unparsed corpus, and only at an extremely low gamma. With these settings, the model is not generally distinguishing between phonologically meaningful gaps and accidental gaps and achieves a poor fit to the data. We hope that future work will incorporate the learning of morphological boundaries and phonotactics into a single model.

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.
- Archangeli, Diana. 1985. Yokuts harmony: Evidence for coplanar representation in nonlinear phonology. *Linguistic Inquiry* 16:335–372.

- Baker, C. L. 1979. Syntactic theory and the projection problem. *Linguistic Inquiry* 10:533–581.
- Bates, Douglas, Martin Maechler, Ben Bolker, and Steven Walker. 2014. *lme4: Linear mixed-effects models using S4 classes*. URL <http://CRAN.R-project.org/package=lme4>, R package version 1.7.
- 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.
- Beckman, Jill. 1997. Positional faithfulness, positional neutralization, and Shona vowel harmony. *Phonology* 14:1–46.
- Bennett, William G. 2013. Dissimilation, consonant harmony, and surface correspondence. Doctoral dissertation, Rutgers University, New Brunswick, NJ.
- 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.
- Bowerman, Melissa. 1988. The “no negative evidence” problem: How do children avoid constructing an overly general grammar? In *Explaining language universals*, ed. John A. Hawkins, 73–101. Basil Blackwell.
- Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.
- Clements, George N. 1976. Palatalization: linking or assimilation? In *Papers from CLS 12*, ed. S. S. Mufwene, C. A. Walker, and S. B. Steever, 96–109. Chicago: Chicago Linguistic Society.
- Cser, András. 2010. The -alis/aris-allomorphy revisited. In *Variation and change in morphology: Selected papers from the 13th international morphology meeting, Vienna*, ed. Franz Rainer, Wolfgang U. Dressler, Dieter Kastovsky, and Hans C. Luschuetzky, 33–51. Amsterdam and Philadelphia: John Benjamins.
- Daland, Robert. 2015. Long words in maximum entropy phonotactic grammars. *Phonology* 32:353–383.
- Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis, and Ingrid Norrmann. 2011. Explaining sonority projection effects. *Phonology* 28:197–234.
- De Lucca, M. 1987. *Diccionario Práctico Aymara-Español, Español-Aymara*.
- 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.
- Dresher, B. Elan. 1999. Charting the learning path: Cues to parameter setting. *Linguistic Inquiry* 30:27–67.

- Dresher, B. Elan, and Jonathan Kaye. 1990. A computational learning model for metrical phonology. *Cognition* 34:137–195.
- Gallagher, Gillian. 2016. Asymmetries in the representation of categorical phonotactics. *Language* 92:557–590.
- Gibson, Edward, and Kenneth Wexler. 1994. Triggers. *Linguistic Inquiry* 25:407–454.
- Goldsmith, John. 1976. Autosegmental Phonology. Doctoral Dissertation, MIT, Cambridge, MA. Published by Garland Press, New York, 1979.
- 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.
- Gouskova, Maria, and Gillian Gallagher. 2017. Inducing nonlocal constraints from baseline phonotactics. Ms., New York University. Currently under review.
- Hardman, Martha James. 2001. *Aymara*, volume 35. München: Lincom Europa.
- Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: The early stages. In *Constraints in Phonological Acquisition*, ed. René Kager, Joe Pater, and Wim Zonneveld. Cambridge: Cambridge University Press.
- 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.
- Kastner, Itamar, and Frans Adriaans. 2018. Linguistic Constraints on Statistical Word Segmentation: The Role of Consonants in Arabic and English. *Cognitive Science* 42:494–518.
- MacEachern, Margaret. 1997. Laryngeal Cooccurrence Restrictions. Doctoral Dissertation, UCLA, Los Angeles.
- Martin, Andrew Thomas. 2007. The evolving lexicon. Doctoral dissertation, UCLA, Los Angeles, CA.
- McCarthy, John J. 1979. Formal problems in Semitic phonology and morphology. Doctoral Dissertation, MIT. Published by Garland Press, New York, 1985.
- McCarthy, John J. 1989. Linear order in phonological representation. *Linguistic Inquiry* 20:71–99.
- McQueen, James M. 1998. Segmentation of continuous speech using phonotactics. *Journal of memory and language* 39:21–46.

- Nazarov, Aleksei, and Gaja Jarosz. 2017. Learning Parametric Stress without Domain-Specific Mechanisms. In *Proceedings of the Annual Meetings on Phonology*, ed. Karen Jesney, Charlie O'Hara, Caitlin Smith, and Rachel Walker, volume 4.
- Prince, Alan, and Bruce Tesar. 2004. Learning phonotactic distributions. In *Constraints in Phonological Acquisition*, ed. René Kager, Joe Pater, and Wim Zonneveld, 245–291. Cambridge: Cambridge University Press.
- Pyle, Charles. 1972. On eliminating BMs. In *Proceedings of CLS 8*, ed. Paul Peranteau, Judith Levi, and Gloria C. Phares, 516–532. Chicago, IL: Chicago Linguistic Society.
- R Development Core Team. 2018. R: A language and environment for statistical computing. Vienna, Austria. URL <http://www.R-project.org>.
- Steriade, Donca. 1987. Redundant Values. In *CLS 23: Parasession on Autosegmental and Metrical Phonology*, ed. A. Bosch, B. Need, and E. Schiller, 339–62. Chicago: Chicago Linguistic Society.
- Suomi, K., J.M. McQueen, and A. Cutler. 1997. Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language* 36:422–444.
- Suzuki, Keiichiro. 1998. A Typological Investigation of Dissimilation. Doctoral Dissertation, University of Arizona, Tucson, AZ.
- Trubetzkoy, N. S. 1939. *Grundzüge der Phonologie*. Prague: Travaux du cercle linguistique de Prague 7.
- Wilson, Colin, and Gillian Gallagher. 2018. Accidental gaps and surface-based phonotactic learning: A case study of South Bolivian Quechua. *Linguistic Inquiry* 49:610–623.

corresponding authors:

Gillian Gallagher, ggillian@nyu.edu

Maria Gouskova, maria.gouskova@nyu.edu

Department of Linguistics
10 Washington Place
New York University
New York, NY 10003