Lexicon against Naturalness: Unnatural Gradient Phonotactic Restrictions in Tarma Quechua

Gašper Beguš   Aleksei Nazarov   Anna Björklund   Blas Puente Baldoceda

Abstract

It has been shown in separate studies that phonological grammar can operate probabilistically and in phonetically unnatural directions. This paper examines whether phonological grammar can be both probabilistic and unnatural at the same time. We create a new corpus of Tarma Quechua vocabulary (based both on published and unpublished data), and argue that the unnatural probabilistic phonotactic trends in the Tarma Quechua lexicon are statistically significant and show clear signs of productivity, with evidence from loanword phonology and from morphophonological alternations. We also perform an acoustic analysis of existing recordings to confirm the phonetic status of the the unnatural gradient restriction. To our knowledge, this is the first report of a fully unnatural gradient phonotactic restriction on segmental structure. The existence of unnatural probabilistic phonology has broad theoretical consequences because it requires probabilistic approaches to phonology to derive unnatural patterns. While weighted constraint approaches are more powerful than non-probabilistic approaches, we argue that models without unnatural constraints are not sufficient for deriving Tarma Quechua data. We propose a new framework for evaluating phonological analyses using the goodness of fit measurement. This quantitative evaluation approach supports analyses with some, but not all unnatural constraints. We argue that the proposed approach can serve for evaluating other models and approaches.

Keywords: naturalness in phonology, analysis evaluation, goodness of fit, gradient phenomena, phonotactics, maximum entropy grammar, indexed constraints

1 Introduction

The question of how phonetic naturalness shapes language is relevant both for theories of synchronic grammar (Chomsky & Halle, 1968; Stampe, 1973; Hellberg, 1978; Donegan & Stampe, 1979; Anderson, 1981; Westbury & Keating, 1986; Archangeli & Pulleyblank, 1994; Hayes, 1999; Wilson, 2006; Carpenter, 2010; Becker et al., 2011; White, 2013; Hayes & White, 2013; Beguš, 2018) and for a discussion on how phonetic, cognitive, and historical factors influence phonological patterns in the world's languages (Moreton, 2008; Blevins, 2004; Beguš, 2019, 2020, 2022). We present here a case of phonotactic restrictions on segmental features (specifically, the feature [± voice]) that are both lexically gradient (Coetzee & Pater, 2008) and phonetically unnatural (Buckley, 2000; Yu, 2004; Coetzee & Pretorius, 2010). Such cases are interesting because they suggest that these phonetically arbitrary or unnatural restrictions extend not only to categorical bans of certain sequences (Coetzee & Pater, 2008), but even to manipulating the relative frequencies of sequences in the lexicon.

1.1 Background

Two questions have received increased attention in the phonological literature: (i) how to represent gradient phonotactic restrictions in the grammar (Frisch et al., 2004; Anttila, 2008; Coetzee &

Pater, 2008; Wilson & Obdeyn, 2009), and (ii) whether and how to represent unnatural processes in the grammar (Hayes, 1999; Hyman, 2001; Blevins, 2004, 2008; Wilson, 2006; Hale & Reiss, 2008; Samuels, 2009; Carpenter, 2010; Coetzee & Pretorius, 2010; Becker et al., 2011; White, 2013; Hayes & White, 2013). To our knowledge, however, there exists no systematic treatment of the intersection of these two topics: unnatural gradient phonotactics, i.e. phonotactic restrictions that, given a particular environment, target a single (segmental) feature and gradiently favor the value of this feature that is unnatural in that environment.

Phonetically unmotivated phonotactics are primarily discussed in light of the "surfeit of the stimulus" problem: there exist unmotivated statistically significant trends in the lexicon, but speakers often fail to generalize these unmotivated trends to nonce words (see, for instance, Becker et al. 2011, 2012; Hayes & White 2013). Hayes & White (2013), for example, identify ten potential "unnatural" phonotactic restrictions based on systematic gaps in the English lexicon and test in a behavioral experiment whether those restrictions are internalized by native speakers. Using Hayes & Wilson (2008)'s Phonotactic Learner to identify systematic gaps, they find that, for instance, the English lexicon features a restriction against /ʒ/ before a stressed vowel followed by an obstruent (*[+continuant, +voice, −anterior] [+stress] [−son] in constraint formalism; Hayes & White 2013), yet learners did not generalize this to nonce words as strongly as they generalize phonetically motivated constraints. Most studies on this problem agree that speakers do not generalize unmotivated phonotactic restrictions to nonce words, with the suggestion that they might not be represented in grammar (Becker et al., 2011, 2012; Hayes & White, 2013; Becker et al., 2017; Wilson & Gallagher, 2018). Other studies (for instance, Hayes 2009) suggest instead that unmotivated processes can actually be generalized to nonce words, but to a lesser extent than natural processes. Some studies also find no differences between opposing patterns (Pycha et al., 2003).

Beside its unnaturalness, the other aspect of our data that bears implications for phonological theory is the fact that restrictions are *gradient* rather than categorical. The necessity of encoding gradient phonotactics in the grammar has recently been motivated on the basis of gradient lexicalized co-occurrence restrictions (Berkley, 2000; Frisch et al., 2004; Anttila, 2008; Coetzee & Pater, 2008; Wilson & Obdeyn, 2009; Zuraw, 2010) as well as other processes (Martin, 2007). Coetzee & Pater (2008), for example, discuss two cases of homorganic consonant co-occurrence restrictions: Arabic and Muna. In both cases, an Obligatory Contour Principle constraint (Leben 1973, McCarthy 1986) creates a pressure that is strong, but fails to reach categorical status because of lexical variation: homorganic consonants are dispreferred within the same root, but not completely ruled out. This dispreference, however, operates in the *natural* direction: a restriction against the co-occurrence of homorganic consonants is strongly motivated by both articulatory (Garrett & Johnson, 2013) and perceptual factors (Gallagher, 2010).

As we will argue in Section 1.2, most cases of what literature thus far calls "unnatural" phonotactic restrictions in the literature are in fact either phonetically motivated or phonetically unmotivated, including the restrictions that were tested in the *surfeit of stimulus* experiments. To distinguish our cases from these, we will adopt a new subdivision of naturalness (Beguš, 2019) and focus on gradient processes that are not only phonetically unmotivated, but that operate against universal, phonetically motivated tendencies.

1.2 On naturalness

Traditionally, the naturalness of phonological processes has been conceptualized as a binary split between natural and unnatural processes (see, for instance, Kiparsky 2006, 2008). Natural processes are phonetically well-motivated and typologically frequent. Their high typological frequency is driven by the existence of phonetic precursors, which Yu (2013, 398) defines as "systematic contextually-induced variations that parallel some cross-linguistically recurring sound patterns". Unnatural processes, on the other hand, are typologically rare or non-existent and lack phonetic precursors.

However, such a division is insufficient, as it fails to identify an important distinguishing feature within the "unnatural" group itself. All processes in the group traditionally called "unnatural" indeed lack phonetic motivation or phonetic precursors. However, a subset of these processes not only lacks phonetic motivation, but also operates against universal phonetic tendencies — that is, against phonetic precursors responsible for a large number of (natural) phonological processes across many languages. Beguš (2019) proposes a new division of phonological processes with respect to naturalness which we follow in this paper.[1]

Consistent with the literature, natural processes are defined as having clear articulatory or perceptual motivations (phonetic precursors; Yu 2013) and being typologically prevalent. Unmotivated and unnatural processes are similar in that they both lack articulatory/perceptual motivation, i.e., they have no phonetic precursors. However, in contrast to unnatural processes, unmotivated processes have no universal phonetic tendencies that militate against them — see the examples in (2), taken from Blevins (2008). On the other hand, (true) unnatural processes not only lack articulatory/perceptual motivation, but also operate against universal phonetic tendencies. In other words, if A → B / X is a universal phonetic tendency, B → A / X is an unnatural process.

To be precise, a *universal phonetic tendency* is a phonetic pressure that has a clear articulatory or perceptual motivation, and operates passively (without active articulatory control) across different languages (Beguš, 2019).

(1) *Universal Phonetic Tendency* (UPT) (Beguš, 2019):
"UPTs are phonetic pressures motivated by articulatory or perceptual mechanisms (for an overview of phonetic mechanisms, see Garrett & Johnson 2013) that passively operate in speech production cross-linguistically and result in typologically common phonological processes."

For example, post-nasal voicing, intervocalic voicing, and word-final devoicing are universal phonetic tendencies: they have clear articulatory motivations, are amply attested cross-linguistically as typologically common phonological alternations, and operate passively even in languages in which such processes are not among the synchronic alternations (for discussion, see Section 2).

The literature on (un)naturalness has primarily focused on unmotivated processes. Most of the 26 listed examples in the survey of "unnatural" alternations in Blevins (2008) and the ten "unnatural" gradient phonotactic restrictions in Hayes & White (2013) are unmotivated rather than unnatural, according to the definition in Beguš (2019). Consider, for example, the following processes from Blevins (2008) and Hayes & White (2013).

(2) *Some processes labeled as "unnatural" in Blevins (2008) and Hayes and White (2013)*

a. /p/ → [s] / __i
b. /i/ → [u] / d__
c. $*\begin{bmatrix} +COR \\ +cont \\ -strid \end{bmatrix} \begin{bmatrix} -stress \\ +round \end{bmatrix}$ ("No [θ,ð] before stressless rounded vowels")

---

[1] Similar, but slightly different distinctions have been proposed before: see Morley (2014). For Morley (2014), for example, "anti-natural" processes need to violate implicational universals and they are defined as "unattested patterns that do not conform with posited language universals".

d. $*\begin{bmatrix}+cont\\+voice\\-ant\end{bmatrix}\begin{bmatrix}+stress\end{bmatrix}\begin{bmatrix}-son\end{bmatrix}$ (“No [ʒ] before stressed vowel + obstruent”)

All these processes lack phonetic motivation, but they do not operate against universal phonetic tendencies. In other words, it is not the case that [θ, ð] in the context before unstressed rounded vowels are universally preferred: there exists no cross-linguistic passive phonetic tendency that would prefer fricatives before unstressed rounded vowels.

Some gradient restrictions that operate in the unnatural direction have recently been reported in light of the surfeit of the stimulus experiments mentioned in Section 1.2, including (i) a higher rate of alternation in monosyllabic compared to polysyllabic words (Becker et al., 2012), and (ii) a stronger preference for antepenultimate stress in LLL words compared to HLL words Garcia (2017). Strictly speaking, the first restriction does not contradict any universal phonetic tendency (perhaps just a universal phonological tendency). The second restriction targets non-segmental features and while the restriction is significant, it is also very subtle in magnitude (see Garcia (2017)).

Few truly unnatural active alternations are reported; there is additional debate as to whether unnatural processes are, in fact, possible as productive synchronic alternations. The most compelling case of an unnatural synchronic alternation is post-nasal devoicing (PND) in Tswana and Shekgalagari (Hyman, 2001). PND qualifies as unnatural according to all criteria laid out in this section: it operates against the well-motivated, typologically common, and passive phonetic tendency of post-nasal voicing. Coetzee & Pretorius (2010) have shown that PND is phonetically real and fully productive: speakers generalize it to nonce words (although it does not appear in all varieties and it might have alternative analyses; Gouskova et al. 2011; Downing & Hamann 2021). On the other hand, one of the most convincing cases against unnatural alternations in synchronic grammars is that of final (de)voicing. While final devoicing is a highly common and phonetically motivated process Steriade 1997; Iverson & Salmons 2011 its unnatural counterpart — final voicing — is not attested in any language as a synchronic process (Kiparsky 2006, 2008; for a possible exception, see Yu 2004; de Lacy 2002; Blevins et al. 2020). This has led some scholars to conclude that final voicing is an impossible synchronic alternation (Kiparsky, 2006, 2008). The question of whether fully productive synchronic alternations can be unnatural remains open.

1.3 Aims

To our knowledge, no systematic treatment of unnatural gradient phonotactic restrictions exists in the literature. This paper presents a novel analysis of the lexicon of the Tarma dialect of Quechua (data from Adelaar 1977; Puente Baldoceda 1977 and novel unpublished data) which has a highly unnatural trend in the lexicon that targets a single segmental laryngeal feature [±voice] at a considerably greater magnitude than any cases of unnatural gradient restriction reported thus far. We show that these trends run counter to specific universal phonetic tendencies, and argue that the trends are statistically significant, phonetically real, and morphophonologically productive. Based on these findings, we interpret unnatural trends as a gradient phonotactic restriction (in line with Coetzee & Pater 2008 and others).

The existence of unnatural gradient phonotactic restrictions has several implications. One of the most consequential debates in phonology is to what degree it is influenced by phonetics and whether it can accommodate unnatural processes (Hyman, 2001; Hayes & Steriade, 2004; Blevins, 2004; Kiparsky, 2006, 2008; Reiss, 2018; Beguš, 2022). Our data speaks to this debate on two fronts: first, we argue that phonotactic restrictions can extend in phonetically unnatural directions. Second, phonological processes are not only categorical, but often operate probabilistically. Our

	[+voice] preferred	**[+voice] dispreferred**
a. Intervocalic voicing	between two vowels	
b. Postnasal voicing	after [+nasal] consonant	
c. Voicing agreement	adjacent to [+voice] consonant	adjacent to [-voice] consonant

Table 1: Some universal phonetic tendencies for [+voice].

presented case suggests that not only categorical, but also gradient phonological processes, can operate in the unnatural direction.

The analysis in this paper raises a theoretical question for the theory of Markedness (see, e.g., de Lacy 2002; de Jong 2004; de Lacy 2006; de Lacy & Kingston 2013) in weighted-constraint theories like Harmonic Grammar or MaxEnt (Goldwater & Johnson, 2003; Coetzee & Pater, 2008, 2011; Pater, 2008, 2009; Albright, 2009; Potts et al., 2010). While weighted constraint models are able to derive gradient phonotactics (see, e.g., Coetzee & Pater 2008), we show that, without unnatural Markedness constraints, they are unable to derive systems in which the unnatural element in that context is more frequent than the natural element in a given context. We call this effect the "Natural Gradience Bias". Since the cases presented here are of this exact nature, this opens up the possibility that Markedness constraints might, after all, be able to counter universal phonetic tendencies (cf. Hayes & White 2013 and other work in that direction).

We also present a new technique to quantitatively evaluate phonological analyses by comparing the goodness of fit of models with different markedness constraints. We model Tarma Quechua data with natural-only, unnatural-only and both natural and unnatural constraints as well as with indexed constraints, then compare the goodness of fit of various models. Such simulations, as well as the proposed evaluation technique, bring several advantages. Firstly, we can test whether phonetically unnatural gradient restrictions can be derived with natural-only constraints. Secondly, we can quantitatively evaluate different analyses and argue in favor of unnatural and lexically indexed constraints based on a quantitative metric. Finally, the technique also facilitates interpretability of MaxEnt models: we provide explanations for various predictions made by the MaxEnt simulations.

2 Universal tendencies for voicing

The case of unnatural phonotactics presented in this paper targets a single phonological feature: [±voice]. Alternations and phonotactic restrictions targeting this feature are among the most well-studied phenomena in phonology and we have a substantial body of research on the typology as well as the phonetics of voicing cross-linguistically. The attention that voicing has received in the literature helps us determine in which environments the voice feature is universally (dis)preferred, i.e. which processes targeting [±voice] are "unnatural" according to the definition above. Table 1 summarizes these tendencies.

A body of research has established that voicing is universally articulatorily dispreferred in word-initial and word-final or coda positions (Westbury & Keating, 1986; Iverson & Salmons, 2011); it is universally preferred intervocalically (Westbury & Keating, 1986; Davidson, 2016) and after a nasal (Rothenberg, 1968; Kent & Moll, 1969; Ohala & Ohala, 1993; Ohala, 1983; Hayes & Stivers, 2000). Voicing is also dispreferred in clusters and geminates: the longer the closure, the more antagonistic articulatory forces are to voicing. Finally, the voicing feature is also universally dispreferred before another obstruent with a different value in voicing: clusters that disagree in [±voice] are universally dispreferred (Myers, 2010).

Tarma Quechua phonotactic restrictions violate naturalness in two main contexts: post-nasally and in consonant clusters. Post-nasal voicing fulfills all three requirements for a universal phonetic tendency from section 1.2. To begin, it is a common productive phonological alternation and sound change (Kümmel, 2007): as reported in Hayes & Stivers (2000), Locke (1983)'s study identifies 15 out of 197 (or 8%) languages surveyed with post-nasal voicing as a synchronic alternation. The articulatory explanation for post-nasal voicing is described in detail in Hayes & Stivers (2000) and Coetzee & Pretorius (2010), building on previous work by Rothenberg (1968), Kent & Moll (1969), Ohala (1983), Ohala (1993), and others. Two basic factors are identified: in the transition from the nasal stop into the oral stop of nasal-plosive clusters, the velum has to rise from its lowest position to a complete closure. During this transition, air can still escape through the nasal cavity ("nasal leakage"), which makes it more difficult to achieve the increase in supraglottal pressure necessary to cease voicing (cf. Coetzee & Pretorius 2010). Secondly, as the velum rises, the volume of the oral cavity increases. This also promotes voicing, as the increased oral cavity volume translates to more time where voicing is possible (Hayes & Stivers, 2000; Coetzee & Pretorius, 2010). Post-nasal voicing is also well-motivated from a perceptual perspective: cues for voicelessness, such as the release burst of a stop, or cues for voicing, such as low frequency energy, are reduced in post-nasal position (Coetzee & Pretorius 2010, p. 405). Finally, voicing is a passive phonetic tendency in post-nasal position, i.e. stops universally feature more voicing into closure post-nasally than in other positions (Hayes & Stivers, 2000; Davidson, 2016).

The second universal phonetic tendency that Tarma Quechua's phonotactics operate against is voicing agreement in obstruent clusters, which, again, fulfills all three conditions mentioned above. Clusters that agree in voicing are typologically very common and voicing assimilation is one of the most common processes cross-linguistically (Myers, 2010). Myers (2010) lists at least 28 such languages; this list is neither exhaustive, nor does it result from a survey. The phonetic motivation is straightforward: laryngeal features "overlap and blend" in obstruent clusters, which results in the passive voicing of preceding voiceless stops due to laryngeal coarticulation Myers (2010, 164ff). Perceptual factors have also been proposed to promote the agreement of laryngeal features (Myers, 2010). Finally, agreement in laryngeal features fulfills the third condition of a universal phonetic tendency: voicing before another voiced stop is a passive phonetic tendency (Barry & Teifour, 1999) and is attested in languages in which voicing assimilation is not a complete synchronic process. As Myers (2010) claims, "[a]coustic studies have shown that there is a longer voiced interval in an obstruent before a voiced consonant than before a voiceless consonant in English (Haggard, 1978; Docherty, 1992; Smith, 1997; Jansen, 2004), French (Snoeren et al., 1992), and Syrian Arabic (Barry & Teifour 1999; Myers 2010, p. 164).

3 Unnatural trends in the lexicon

This section shows that the identified trends in the lexicons that operate against the two universal phonetic tendencies are statistically significant, phonetically real, and point to evidence in favor of the productivity of these processes.

3.1 Tarma Quechua

Tarma Quechua is a dialect of Quechua (Quechua I) spoken in the Tarma district of the Junín province of Peru. The number of speakers is difficult to establish, as the dialect is rapidly being replaced by Spanish (Adelaar, 1977). Adelaar (1977) and Puente Baldoceda (1977) report approximately 30,000-40,000 inhabitants of the Tarma district and an additional 3,500 inhabitants of La Unión Leticia. However, the number of speakers of the dialect with the particular unnatural phonotactics that are of interest here is much smaller and difficult to estimate. The unnatural phonotactic restriction presented here is also reported in the dialect of Paccho (Adelaar & Muysken, 2004). Because there are no descriptions or recordings of Paccho Quechua available, we leave this dialect out of our discussion.

3.1.1 Stop voicing

The Quechua dialect continuum almost uniformly has only voiceless stops in native vocabulary (Adelaar & Muysken, 2004). Adelaar (1977) and Puente Baldoceda (1977) report that some of these voiceless stops have become voiced in Tarma Quechua (henceforth: TQ), or more precisely, the Quechua dialects spoken in Tarma, Huaricolca, Palcamayo, and La Unión Leticia. Voiceless velar and labial stops (to the exclusion of alveolars) are reported to undergo voicing in intervocalic and post-consonantal positions, but not after a nasal consonant. Adelaar (1977) and Puente Baldoceda (1977) note that voicing does not apply categorically (and that it does not apply post-nasally), but no further analyses on the lexicon are performed. Below we present results of a statistical analysis of the TQ lexicon that reveals a highly unnatural lexical trend. We show that, in addition to limitations on voicing after nasals, the relative rates of voicing in different environments, including intervocalic and post-consonantal positions, contradict the universal phonetic tendencies from Section 2. This paper, to our knowledge, is the first report of this unnatural distribution in TQ.

3.2 Data

For the purpose of the analysis, we collected all tokens of prevocalic labial and velar stops from the vocabulary list in Adelaar (1977) and an unpublished lexicon by Puente Baldoceda. Altogether 2,097 entries were thus collected. The entries were annotated for place of articulation (labial or velar), voicing, part of speech, loanword status, derived status, left and right context, source dictionary, and syllable position. For details on data collection and annotation, see Section A.1 in the Supplementary Materials.

3.3 Statistical analysis

3.3.1 Individual segments

The raw data analysis of non-loan non-derived vocabulary in Table 2 reveals a surprising trend: voicing surfaces very rarely post-nasally (11.7%), in almost half of the lexicon intervocalically (43.9%), and almost always post-consonantally, including in positions after a voiceless fricative (78.0%) or voiceless stop (90.6%).[2] Raw counts suggest that voiced stops are least frequent word-initially and postnasally, and most frequent after voiceless obstruents. To test the statistical significance of this trend, we fit a logistic regression model to the data with the R statistical software (R Core Team) using the *glm()* function. The dependent variable was binary: presence or absence of voicing (voicing coded as success); the independent variables were PLACE of articulation, and LEFT CONDITION (i.e., the phonological class of the immediately preceding segment; additionally, when Left Condition = V, only intervocalic segments are selected). The model also includes the PART OF SPEECH predictor (with four levels; nouns, verbs, dual noun/verbs, and other) to control for potential effects of parts of speech on the voicing pattern and the DICTIONARY source predictor (to test whether the described voicing patterns differ across the two dictionaries).

The initial full model was fit with all predictors and their interactions. The best fitting model was chosen with the step-wise backwards model selection technique: higher order interactions

---

[2] Unless noted otherwise, we will henceforth abbreviate phonological classes as follows: T – voiceless stop, D – voiced stop, S – voiceless fricative, N – nasal, R – non-nasal sonorant, including glides, V – vowel.

	#__	N__	V__V	R__	S__	T__
voiced	14	11	158	82	71	29
voiceless	417	83	202	28	20	3
% voiced	3.2	11.7	43.9	74.5	78.0	90.6

Table 2: Voiced vs. voiceless labial and velar stops in Tarma Quechua native vocabulary across contexts.

were removed step-wise from a full model. If the Akaike Information Criterion (AIC) determined an interaction or predictor does not improve fit significantly, they were removed until all predictors in the model significantly improved the fit. The best-fitting model includes PLACE, LEFT CONDITION, and DICTIONARY as predictors and two interactions: LEFTCOND:DICTIONARY and PLACE:DICTIONARY. The PART OF SPEECH predictor is not significant according to the AIC test.

The DICTIONARY predictor has two levels: Adelaar includes all data entries that are in Adelaar or in both dictionaries; Baldoceda includes those entries that appear in the Baldoceda dictionary only and are absent from Adelaar. This predictor thus indicates whether either data source deviates significantly from the other. Figure 7 in Section A.2 in the Supplementary Materials shows that the major trends in the Adelaar's dictionary hold in those lexical items in Baldoceda that are absent from Adelaar. This suggests that the voicing pattern is independently confirmed in two sources. However, because the predictor DICTIONARY is not linguistically justifiable (lexical items that are absent from Adelaar 1977, but present in Baldoceda have no linguistic function; their only function is to test whether the trends persist across dictionaries), we exclude this predictor from the final model.

The final model includes PLACE (treatment-coded with two levels: labial and velar, with labial as the reference level)[3] and LEFT CONDITION (treatment-coded with five levels: initial, post-nasal, intervocalic, post-sonorant, and post-fricative and post-stop, with intervocalic as the reference level). While their interaction is not warranted by AIC, we leave it in the model. The significance of the estimates of interest does not differ between the best-fitting model (which includes the DICTIONARY predictor) and the final model.

The estimates of the final model in Figure 1 and Table 3 illustrate the unnatural trends in the Tarma Quechua lexicon. In the intervocalic position (V_V), voice contrasts fully: voiced and voiceless stops are equally frequent in this position ($\beta = -0.18, z = -1.62, p = 0.11$). Voiced stops are significantly less frequent word-initially ($\beta = -3.40, z = -9.79, p < 0.0000$) and postnasally than intervocalically ($\beta = -1.96, z = -4.61, p < 0.0000$). On the other hands, voiced stops are significantly more frequent after sonorants and glides ($\beta = 1.52, z = 5.20, p < 0.0000$), after voiceless fricatives ($\beta = 1.67, z = 5.16, p < 0.0000$), and after voiceless stops ($\beta = 2.49, z = 3.83, p = 0.0001$) than they are intervocalically. No overdispersion was detected in the models (ratio = 1.01; tested with *overdisp_fun()* in Bolker 2019).

These trends are significant even if we include loanwords and derived words in the analysis. Data with the entire collected vocabulary including loanwords and derived words (total N=1,793) were fit to the same model with LEFT CONDITION, PLACE and their interaction as predictors. Estimates in Table 16 suggest all differences remain significant in this model.[4]

Furthermore, if we add LOANWORD status as a predictor to our model (including loan and derived vocabulary), the fit improves significantly, but it is not clear whether loanword status as a predictor is justifiable in a cognitive model. It is likely that speakers are unaware of loanword

---

[3] For a discussion on contrast coding, see Brehm & Alday (2022).

[4] The interaction is significant in this model.

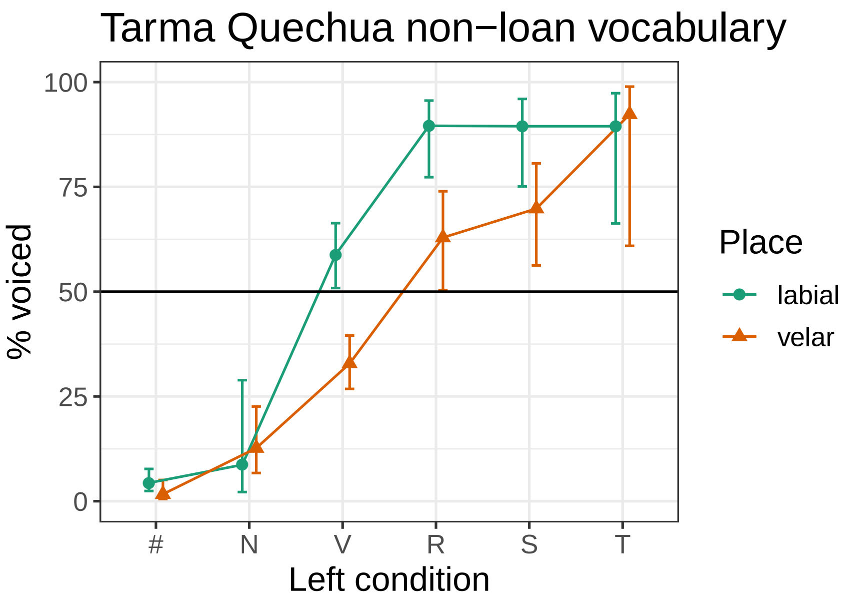


Figure 1: Percentage of voiced stops according to position based on a model in Table 3. All bars in figures represent 95% confidence intervals based on model estimates.

status for many lexical items; if they are, the distribution of [*pm*voice] for loanwords ceases to be of interest (in this case speakers may use two different grammars that govern native and loanword phonologies; see, e.g., Itô & Mester (2002); Itô & Mester (2003). The data were fit to a model with three independent variables: Left Condition, Place of articulation, and Loanword status (sum-coded with native words as the reference level). Two two-way interactions are significant (Left Condition:Place and Left Condition:Loan) and are included in the model. The significance of all main effects remains the same as before (at means of other predictors), except that the V__V vs. S__ and V__V vs. T__ cease to be significant as a main effect, likely due to data scarcity. There is only one loanword with labials or velars in post-stop position, and only twelve in post-fricative position. Other frequency differences remain significant: there is still more voicing intervocalically than post-nasally.

If we isolate loanwords from the native vocabulary, we do not observe any unnatural patterns, which is not surprising as the donor language, Spanish, does not feature any of the unnatural patterns in TQ. In a model that excludes derived words, but includes loanwords with three predictors (Left Condition, Place, and Loan with all interactions), the loanword vocabulary does not reflect the unnatural pattern from the non-loan vocabulary (see Figure 8). Adding loanword status as a predictor also introduces a problem of data scarcity, since there are only a few loanwords with labials or velars in post-obstruent position. As will be shown below, however, the unnatural voicing pattern in TQ does apply to a subset of loanwords.

To test the effect of syllable structure, we fit a model with Left Condition and Syllable structure as predictors with their interaction (on non-loan non-derived vocabulary). Adding Syllable structure introduces highly sparse data in some conditions, which is why it is not added to the main model and which is why we do not include Place as a predictor here. We divide data into four groups according to the syllable structure: onset of the ultimate syllable in disyllabic words (the most common position given TQ root structure), onset of the ultimate syllable in trisyllabic words, onset of the penultimate syllable in trisyllabic words, and all other

	Estimate	Std. Error	z value	Pr(>|z|)
(Intcpt)=V	-0.1791	0.1105	-1.62	0.1051
#	-3.4002	0.3474	-9.79	0.0000
N	-1.9615	0.4254	-4.61	0.0000
R	1.5191	0.2921	5.20	0.0000
S	1.6683	0.3232	5.16	0.0000
T	2.4916	0.6502	3.83	0.0001
Place1	0.5358	0.1105	4.85	0.0000
#:Place1	-0.0433	0.3474	-0.12	0.9007
N:Place1	-0.7466	0.4254	-1.76	0.0792
R:Place1	0.2760	0.2921	0.95	0.3447
S:Place1	0.1150	0.3232	0.36	0.7219
T:Place1	-0.7082	0.6502	-1.09	0.2760

Table 3: Logistic regression model with PLACE (sum-coded) and LEFT CONDITION and their interaction as predictors (non-loan non-derived vocabulary)

syllabic positions (elsewhere). The estimates of the model in Figure 9 suggest that in the ultimate syllable of disyllabic words, all trends are as observed in the full model. Occurrences of labial and velar stops in trisyllabic words are rather rare, and the effects are not significant. Raw data, however, suggests, that the same tendencies do also operate in trisyllabic words. For example, in the postnasal position in trisyllabic words, 13 (76.5%) stops are voiceless and 4 are voiced (23.5%). In the position after a voiceless stop (T), 2 stops are voiced (66.7%) and 1 is voiceless (33.3%). A notable deviation from the general trend is that in trisyllabic words, stops in a position after a voiceless fricative are voiceless in 5 cases and voiced in 1 case. In sum, the observed unnatural differences in voicing are confirmed when they all appear in the most common position given TQ's syllable structure: the ultimate syllable of disyllabic words. In trisyllabic words, the data are sparse, but at least the distinction in voicing between post-nasal and post-stop positions operate in the expected unnatural direction.

As shown in this analysis, [+voice] in labial and velar stops is significantly less frequent word-initially and post-nasally compared to in intervocalic position. [+voice] is significantly more frequent in post-sonorant and post-obstruent position compared to intervocalic position in the TQ native vocabulary. These trends hold across parts of speech, syllable structures, and source dictionaries.

As argued in Section 2, post-nasal and intervocalic positions universally prefer voicing, while voiced stops after voiceless obstruents are universally dispreferred. The fact that TQ exhibits less voicing post-nasally and intervocalically than after voiceless obstruents (labial and velar stops never occur after a voiced obstruent in our data)[5] is thus highly unnatural. TQ voicing thus operates in a direction opposite to two universal phonetic tendencies: it operates more frequently where it is universally dispreferred (post-consonantally) and less frequently where it is universally preferred (post-nasally and intervocalically). These findings are summarized in the table in 4.

---

[5] The analysis shows post-nasal < intervocalic and intervocalic < post-obstruent, from which post-nasal < post-obstruent can be derived by transitivity.

1^{st} member	**2^{nd} member**	
	Labial	**Velar**
t	lutbi	mutgi
ʧ	/	aʧga
ʈ͡ʂ	aʈ͡ʂba	maʈ͡ʂga
k	takba	/
s	ʧasbu	ʈ͡ʂasgi
ʃ	kaʃbi	iʃgi
x	saxbi	manexax-gunas
l	ʧilbi	ʧilgi
r	karba	argu
j	ajba	ajga
w	kawbu	awgis

Table 5: Obstruent clusters in TQ (from Adelaar 1977).

Universal tendencies for [+voice]	Observed significant trends in TQ
T__ < V__V	V__V < T__
T__ < N__	N__ < V__V < T__

Table 4: Unnatural distribution of [+voice]

3.3.2 Disagreement of voicing in clusters

Another locus of gradient unnaturalness emerges in TQ if we look into the within-context distribution of voicing in consonant clusters involving obstruents: obstruent clusters that agree in voicing are gradiently dispreferred in TQ — clusters that disagree in voicing are significantly more frequent.

In TQ, labial and velar stops surface as voiced in non-nasal post-consonantal position (Table 2). The following consonants are attested as triggering voicing: [t, ʧ, ʈ͡ʂ, k, s, ʃ, x, l, lʲ, r, j, w]. The list includes voiceless fricatives, affricates, and even voiceless stops. The following clusters of two stops are attested: [kb, tb, tg]. Table 5 presents examples of clusters that disagree in voicing after each consonant (data from Adelaar 1977).

A statistical analysis of this trend reveals that obstruent clusters that disagree in voicing are much more frequent than clusters that agree in voicing if the second consonant is either a labial or a velar stop. Table 6 shows the number of occurrences of obstruent clusters (both stops and fricatives; for the purposes of this particular analysis, we mark all voiceless obstruents with T, and all voiced obstruents with D) in which the first element is an obstruent and the second element is a labial or a velar stop in the non-derived non-loan TQ vocabulary. To test the statistical significance of this distribution, the data were fit to a logistic regression model with voicing as the dependent variable, Place of articulation and Cluster position (first vs. second) without the interaction (which is not significant). In the second position in obstruent clusters (after the first obstruent which in all but two cases are voiceless), voicing is significantly more frequent than non-voicing ($\beta = 2.25, z = 5.21, p < 0.00001$). Voicing is significantly less frequent when the obstruent is in the first position compared to the second position ($\beta = -5.48, z = -6.91, p < 0.00001$). Tested differently, the proportion of clusters that disagree in voicing is significantly higher than 50% (according to the Exact binomial test; $p < 0.00001$).

TQ thus features a statistically significant trend that restricts clusters agreeing in voicing in

	TT	**TD**	**DT**	**DD**
Count	23	100	2	0
Percent	18.4%	80.0%	1.6%	0%

Table 6: Voice feature in obstruent clusters in which the first element is an obstruent and the second element is a labial or a velar stop in non-derived non-loan TQ vocabulary.

favor of disagreeing clusters. This trend is both gradient and unnatural.

The trend against agreeing obstruent clusters in TQ is unnatural in one additional respect. Table 6 shows a preference for TD clusters, compared to DT clusters — which goes against yet another phonetic tendency. Voicing is articulatorily easier to maintain in the initial parts of closure than it is to initiate voicing after a period of voiceless closure (Ohala & Riordan, 1979; Ohala, 1997). The reason for this articulatory dispreference is straightforward and has been identified as the Aerodynamic Voicing Constraint: airflow and a subglottal-supraglottal pressure difference, necessary for voicing, are sufficient during vowel articulation, but decrease into closure. The reason why voicing is articulatorily difficult to initiate after a period of voiceless closure is that it is difficult to reinstantiate the necessary airflow and pressure difference — once the closure has caused them to decrease — without releasing the stop closure completely. In addition, there is a typological tendency towards respecting the Syllable Contact Law (Vennemann, 1988), which also prefers DT over TD clusters. Finally, decreasing phonation into closure is observed as a passive tendency in several languages (see, for instance, Möbius 2004; Davidson 2016). In other words, voicing has a universal tendency to decrease rather than increase during the closure. The restriction in TQ against DT (decreasing in voicing) clusters in favor of TD clusters (increasing in voicing) is thus unnatural: it operates against the universal phonetic tendency that decreases voicing into closure.

3.4 Phonetics

No previous detailed phonetic analyses of the system of voicing in TQ exist: Adelaar (1977) and Puente Baldoceda (1977) are based on qualitative descriptions of recordings and are not supported by acoustic analyses. Our analysis confirms the phonetic reality of the TQ voice system as described above, making the case for true unnaturalness in the TQ data.

The analyzed recordings were obtained online[6] in .wav format, sampled at 90 kHz[7] with 16-bit quantization and analyzed with Willem Adelaar's permission in the Praat software (Boersma and Weenink 2016). The recordings were made by Willem Adelaar in 1970 in Tarma, in the Junín province of Peru. The informant was a 35 year old male speaker of TQ. The recordings are noisy with considerable reverberation (for a detailed acoustic analysis of reverberation, see Section A.3 in the Supplementary Materials), but the analysis nevertheless reveals important aspects of the unnatural gradient phonotactics and of the phonetic system of TQ in general.

Figure 2 shows four waveforms and spectrograms of two TD clusters: [tb] and [kb]. All four spectrograms clearly show that the initial stop of the cluster is voiceless with almost no phonation into closure and that phonation does not start until the onset of the second stop's closure.

---

[6] Accesible online at: https://corpus1.mpi.nl/ds/asv/?0&0%5C&openpath=node:1483874

[7] Although not specified by the archive from which the audio was retrieved, it can be assumed from their age that the recordings were originally analogue, and therefore have no original sampling frequency.

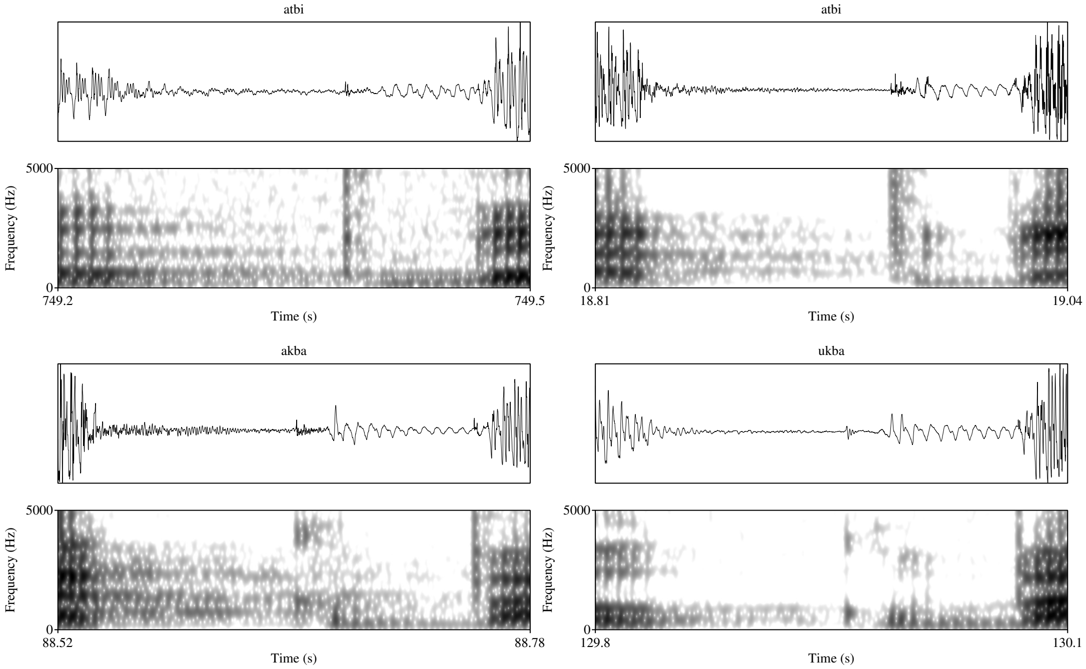


Figure 2: Waveforms and spectrograms of four TD clusters: [atbi], [atbi], [akba], and [ukba].

The same situation holds when the first element is an obstruent other than a stop, such as in the [sb] or [sg] sequences in Figure 3. The lack of any low frequency energy in the fricative portion of the cluster confirms that the first element, [s], is voiceless. Phonation starts at the onset of the stop.

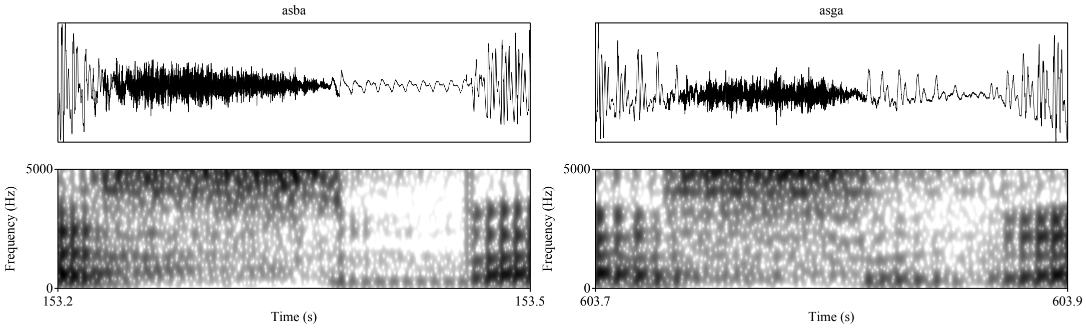


Figure 3: Waveforms and spectrograms of two TD clusters: [sb] and [sg].

The exact realization of voiced stops in clusters is not completely uniform and may vary. The exact distribution is difficult to establish with limited data, but a short transitional vocalic element is occasionally found between the voiceless and voiced obstruent, indicating a smaller degree of gestural overlap (Figure 5).[8] Occasionally, the voiced element surfaces as a fricative in apparent free variation. Figure 4 presents waveforms and spectrograms of phonemic voiced stops that surface as voiced fricatives. Spectrograms clearly show that the manner of articulation of sounds in question is frication. Formants are present throughout the consonantal part, even in cases where formants do

---

[8] Occasionally, the second element is found to surface as voiceless or deleted.

not result from reverberation see Section A.3 in the Supplementary Materials) — after a voiceless stop. Moreover, fricatives feature a gradual increase in amplitude and lack bursts (bursts being a characteristic of stop consonants); both of these characteristics are confirmed in the spectrograms.

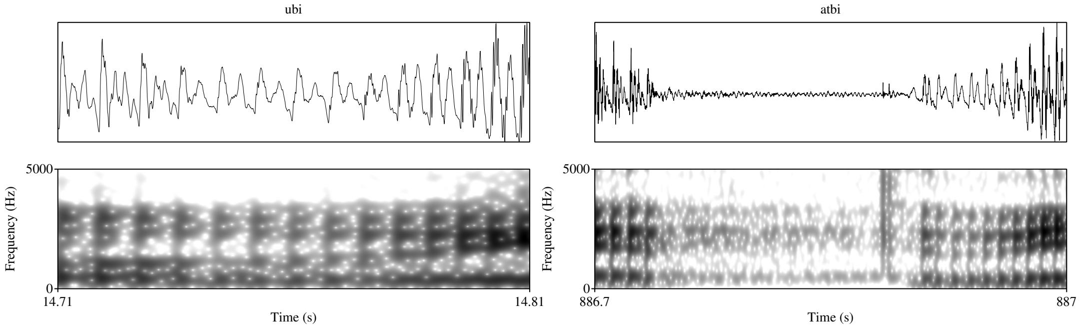


Figure 4: Waveforms and spectrograms of voiced stops that surface as voiced fricatives: intervocalically (left) and after a voiceless stop (right).

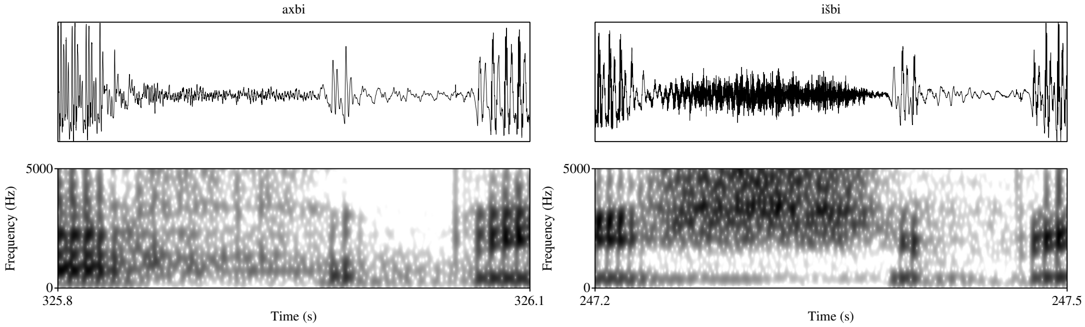


Figure 5: Waveforms and spectrograms of voiced labial stops in post-consonantal position with a short vocalic element between the voiceless and voiced element: [xb] (left) and [ʃb] (right).

After nasals, on the other hand, voiceless stops are the preferred variant, as detailed in Table 3. Figure 6 shows spectrograms with voiceless [p] and [k] after nasals. Also note that voiceless stops in TQ are unaspirated, which means that the phonotactic restriction in fact targets the feature [±voice] rather than the feature [±spread glottis].

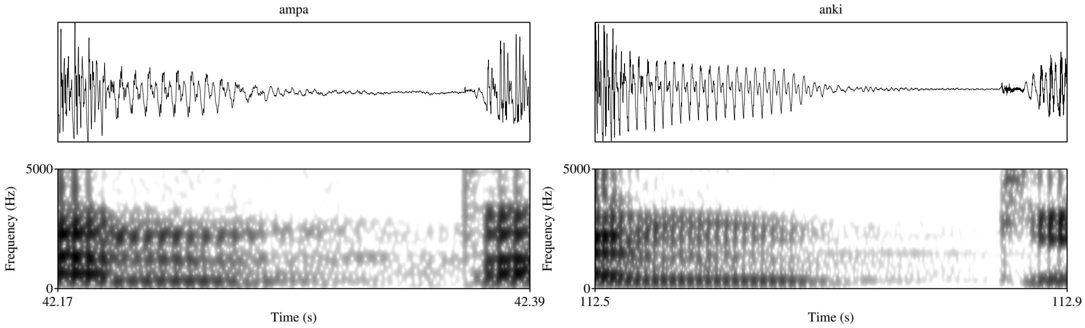


Figure 6: Waveforms and spectrograms of voiceless stops in post-nasal position: [mp] and [ŋk] (right).

3.5 Productivity

Corroborating its status as a phonotactic restriction, there exists evidence that the gradient phonotactic restriction is synchronically active in some morphophonological alternations. Creider (1968) and Adelaar (1977) identify four suffixes with an initial voiced labial stop that feature morphophonemic alternation:

(3) *Alternating suffixes*
  a. -ba/-pa 'genitive'
  b. -bax/-pax 'purposive'
  c. -bita/-pita 'procedentive'
  d. -bis/-pis 'even, too'

The allomorph with voiced initial stops is selected after vowels and non-nasal consonants, including voiceless obstruents; the allomorph with voiceless initial stop is selected after nasals (Creider, 1968). The distribution is illustrated in (4).

(4) a. *Intervocalic*
  wawxi-gi-**ba** wayi-n
  'the house of your brother'
  b. *Post-nasal*
  wayi-n-**pa** pasa-ʃun
  'we're going to walk by way of his house'
  c. *Post-obstruent*
  tamya-ya-n nuqa-nt͡ʃik-**baq**
  'it is raining now for us' (Creider, 1968, p. 12-13)

This process is productive for a subset of suffixes. Other suffixes do not enter the alternation. For example, the highly frequent plural suffix /-guna/ and other suffixes /-bura/, /-gama/, and /-gasga/ have no voiceless allomorphs in post-nasal position (Adelaar, 1977, p. 59). The productivity of this morphophonemic alternation also differs across the dialects. (Adelaar, 1977, p. 59) reports that voiceless allomorphs are required in Vicora Congas, whereas in Huanuquillo the rate of application varies, i.e. is gradient.

The behavior of loanwords provides further evidence for the productivity of unnatural gradient phonotactics. Most Spanish loanwords retain their original voicing. Sporadically, however, voicing of an originally unvoiced segment or devoicing of an originally voiced segment does occur (data from Adelaar 1977).

(5) a. Sp. *cuculi* > kuguliː 'white-winged dove'
  b. Sp. *cotpe* > kutbi 'an animal from the mountains'
  c. Sp. *sauco* > sawgu 'magic tree'
  d. Sp. *vaca* > waːga 'cow'

In two loanwords, a Spanish voiced intervocalic stop devoices to a TQ voiceless stop (data from Adelaar 1977).

(6) a. Sp. *taruga* > taruka 'deer'
  b. Sp. *dios se lo pague* > jusulpaːki 'thank you'

The two loanwords with devoicing of intervocalic stops are especially relevant for the discussion on the productivity of TQ unnatural gradient phonotactic restrictions. Devoicing in TQ [taruka] from Sp. *taruga* cannot be a result of an earlier borrowing, when TQ voicing supposedly did not operate yet. The historical development of TQ involves only voicing of voiceless stops, not devoicing of voiced stops. Regardless of when Spanish *taruga* was borrowed, sound change could not have produced TQ [taruka]. This means that the gradient phonotactic restriction was likely productive and resulted from the law of frequency matching (Hayes et al., 2009): because voiced stops surfaced in approximately half of the lexicon, nativization that matches native vocabulary frequencies is predicted to occasionally voice voiceless stops of the donor language and devoice voiced ones.[9]

4 Grammar learning simulations

The results from the statistical and acoustic analysis of the unnatural trends in TQ in Section 3 suggest that the grammar needs to derive not only phonetically motivated, but also phonetically unnatural gradient processes in phonology. Whether phonetically unnatural constraints are needed in the grammar has been a long-standing discussion in phonology (Hyman, 2001). In this section, we tackle this question by simulating grammars with natural and unnatural weighted constraints (within the MaxEnt framework). We propose a novel quantitative evaluation technique that identifies which natural/unnatural constraints are necessary for deriving gradient phonotactic restrictions. In doing this, we demonstrate the existence of a "natural gradience bias" in Maximum Entropy (MaxEnt) grammars: the presence of only natural constraints cannot motivate all unnatural gradient processes, *pace* Magri & Anttila (2018).[10] Our technique also facilitate interpretation of what various MaxEnt models are learning and where they fail to learn phonotactic generalizations.

In other words, we compare Maximum Entropy (Goldwater & Johnson, 2003; Hayes & Wilson, 2008) learning models with different constraint sets to find how the presence of particular unnatural constraints affect the models success in capturing the data. We also consider a MaxEnt model that includes indexed constraints (Pater, 2000, 2010), which is able to produce an even better fit for the data by accounting for the fact that each word has a uniform pronunciation.

4.1 Setup of MaxEnt Models

All our MaxEnt models are set up with the following assumptions. Every TQ word found in the corpus (in Section 3) is considered to be a separate input. Since we want to model the voicing of labial and velar stops in the grammar, it is assumed that the underlying voicing of all labial and velar stops is underspecified, while the surface candidates feature all possible combinations of voiced and voiceless variants of each of these stops. Consider, for instance, the word ʈ͡ʂimpajgu- 'to (make) kneel [arodillar(se)]': its corresponding input is /ʈ͡ʂimPajKu-/ (with P and K being labial and velar stops unspecified for [voice]) and the output candidates will be [ʈ͡ʂimpajku-, ʈ͡ʂimbajku-, ʈ͡ʂimpajgu-, ʈ͡ʂimbajgu-]. In this setup, there are never any feature mismatches between input and output, which means that faithfulness constraints like IDENT(voice) would have no violations for any of the inputs and surface candidates. Because faithfulness constraints would have contributed

---

[9] This assumes, of course, that these loanwords were not borrowed into TQ via some other Quechuan dialect without this particular voicing pattern after the historical development of this pattern was completed in TQ.

[10] Anttila (2008) alludes to this generalization, but does not provide a more elaborate account. For an earlier treatment of "natural gradience bias", see Beguš & Nazarov (2018). Hayes (2016) argues that in MaxEnt with a restricted CON "[a] harmonically bounded candidate can never receive a higher probability than the candidate that bounds it" and calls this generalization "stochastic harmonic bounding".

Family	**Natural constraints**	**Unnatural constraints**
Context-free	*[+voice], *b, *g	*[-voice], *p, *k
Word-initial	*#[+voice], *#b, *#g	*#[-voice], *#p, *#k
Voicing agreement	Agree(voice),	DisAgree(voice)
	Agree(voice)/[+labial],	DisAgree(voice)/[+labial],
	Agree(voice)/[+dorsal]	DisAgree(voice)/[+dorsal]
Coda condition	CodaCond([+voice]),	CodaCond([-voice]),
	CodaCond([+voice])/[+labial],	CodaCond([-voice])/[+labial],
	CodaCond([+voice])/[+dorsal][11]	CodaCond([-voice])/[+dorsal]
Intervocalic	*VTV, *VpV, *VkV	*VDV, *VbV, *VgV
Postnasal	*NT, *Np, *Nk	*ND, *Nb, *Ng
Post-R	*RT, *Rp, *Rk	*RD, *Rb, *Rg

Table 7: Constraints considered in our MaxEnt models.

nothing to the analysis in this particular case, we only include markedness constraints, making the result a completely phonotactic grammar.

We consider 42 constraints: 21 typologically and phonetically natural ones and 21 unnatural mirror images. These can be split up into 7 families based on the context in which they operate (Table 7).

The natural constraint families can be justified as follows. There is a general bias against voiced obstruents both across and within languages (Ohala, 1983) — hence the constraint *[+voice] or, in short, *D (Lombardi 2001; this setup does assume that sonorants and vowels do not have [voice] contrastively). Voicing is also generally avoided word-initially (Westbury & Keating, 1986), given processes like word-initial strengthening: hence the constraint *#[+voice]. Voicing agreement within clusters is a widely attested phenomenon (Myers, 2010); we use the Agree constraint family (Bakovic 2000), but another way of dealing with agreement could be substituted (e.g., Agreement by Correspondence, Rose & Walker 2004; Share, McCarthy 2011). To account for the fact that marked feature values are preferred in onset position or in a coda assimilated to an onset, Ito (1990) proposed the concept of Coda Condition, later adapted as the Coda Condition constraint family (e.g., Itô & Mester, 1994; Zoll, 1998; McCarthy, 2007, 2008b). Finally, constraints against intervocalic (Westbury & Keating, 1986; Kaplan, 2010; Davidson, 2016) and postnasal (Pater, 1999) voiceless obstruents have been proposed to account for the avoidance of such configurations across languages. The *RT constraint family is an extension of these constraints, based on the general idea that a sonorant prefers to be followed by a voiced rather than voiceless obstruent.

The place-specific versions of these constraints (specific to [+labial] or [+dorsal] consonants) are violated whenever the target obstruent in the constraint definition is labial or velar, respectively. The place-specific CodaCond constraints target codas that are labial or velar. The place-specific Agree constraints are violated whenever at least one obstruent in a cluster is labial (or velar, respectively) and the obstruents disagree in voicing.

The unnatural counterparts of these constraints are obtained by reversing the value of [voice] in the constraint, if said constraint only refers to one of the values (+ or −). As the unnatural counterpart of Agree, we devised the constraint DisAgree, which assigns a violation mark every time that two consecutive obstruents have equal values of [voice]. Its place-specific variants are defined the same as for the Agree family: for instance, DisAgree(voice)/[+labial] is violated once for every obstruent cluster that contains a labial obstruent and in which the obstruents agree in voicing.[12]

With these constraints, we built several models, 12 in total. These comprise 3 basic models: full (all constraints), all-natural, and all-unnatural; 8 blended models, which have a tailored mix of natural and unnatural constraints; and 1 model with indexed constraints, which is considered in Section 4.4.

4.1.1 Basic models

The full model has all 42 constraints in it: the learner was able to weight both natural and unnatural constraints to account for the pattern. This model gives the learner the greatest chance of succeeding and is tested to see what roles the learner assigns to natural vs. unnatural constraints.

The other 10 models (except the model with indexed constraints) contain just 21 constraints each, which is done to test the effect of the presence of natural vs. unnatural constraints without varying the number of constraints, since a greater number of constraints increases a model's chance of accounting for the data (see, e.g., Hastie et al. 2009 for the general principle that adding more parameters to a model increases its chance of better fitting the training data). The all-natural model has all 21 natural constraints, while the all-unnatural model has all 21 natural constraints. The all-natural model provides a baseline of how well a model with only universal, natural constraints fares on the TQ data. The all-unnatural model provides a polar comparison, given that it does not have the opportunity to encode any natural tendencies in the TQ data.

4.1.2 Blended models

Furthermore, for each of the 7 unnatural constraint series in Table 7, a blended model is considered that combines the unnatural constraints from one family with the natural constraints from all other families. For instance, the *ND model contains the unnatural constraints from the Postnasal family, but the natural constraints from all other families. This keeps the number of constraints the same between the models. Each of these models tests whether including the unnatural constraints from one of the families improves the fit of the resulting optimized model to the data compared to the all-natural model. Given the statistical results in Section 3.3, we predict that unnatural constraints from the Postnasal and Voicing agreement families (*ND, DisAgree) will lead to an improved fit to the data, which would show that unnatural constraints from these families are necessary to model the gradient phonotactic pattern. This is what we mean by a "natural gradience bias": if only natural constraints are present, the unnatural gradience in our data cannot be accounted for.

Given the latter prediction, we also trained one last model: the DisAgree & *ND model, which contains unnatural constraints from the Voicing agreement and Postnasal families, but natural constraints from all other families. This model tests whether adding constraints in favor of voicing disagreement and postnasal voicing constitutes an improvement over a model with just voicing disagreement or just postnasal voicing.

An overview of the models is given in Table 8, which lists for every model whether the natural or the unnatural version of a given constraint family is included into that model.

---

[12] For an alternative solution, DisAgree(voice) could be replaced by the combination of OCP(voice) (see Itô & Mester 1986, 2003) and a constraint against two consecutive obstruents sharing the same autosegmental voicing specification. However, we chose the one-constraint option so that grammars with natural and unnatural constraints could have matching numbers of constraints.

Model type	Model name		Context-free constr. natural?	#_	Agree	CodaCond	V_V	N_	R_
Basic	Full	All-natural	✓	✓	✓	✓	✓	✓	✓
		All-unnatural	✗	✗	✗	✗	✗	✗	✗
Blended		*[-voice]	✗	✓	✓	✓	✓	✓	✓
		*#[-voice]	✓	✗	✓	✓	✓	✓	✓
		DisAgree	✓	✓	✗	✓	✓	✓	✓
		CodaCond([-voice])	✓	✓	✓	✗	✓	✓	✓
		*VDV	✓	✓	✓	✓	✗	✓	✓
		*ND	✓	✓	✓	✓	✓	✗	✓
		*RD	✓	✓	✓	✓	✓	✓	✗
		DisAgree & *ND	✓	✓	✗	✓	✓	✗	✓

Table 8: Naturalness of constraint families (columns) in each model (rows).

4.2 Learning setup

All 12 models (including the one with indexed constraints) are trained with Staubs' (2011) MaxEnt training package, which optimizes constraint weights given an array of tableaux with pre-defined OT-style constraints. Since the data are categorical, the objective of the learner is to maximize data log-likelihood (or, in practice, minimize negative data log-likelihood). The models are optimized with the built-in *optim* function in R (R Core Team, 2018) with the L-BGFS-B method (Byrd et al., 1995), L2 optimization with μ=0 and σ=1,000,000, and a maximum number of iterations of 100,000. Since the models have no hidden structure (cf. Tesar & Smolensky 2000), there is no reason to assume that different runs and different initializations of this algorithm will yield different optima. Nevertheless, each model is run 5 times with all constraint weights initialized at 0. No differences were observed among the 5 runs for any of the models, so the results presented in the next subsection will not refer to distinct runs of the same model.

4.3 Results

The MaxEnt grammars optimized by this method are evaluated quantitatively in terms of the (training) data log-likelihood, and qualitatively in terms of their performance on a test dataset and in terms of their constraint weights. The test dataset consists of labial and velar stops in all relevant environments, as shown in Table 9. Like in the training dataset, surface candidates differ in the voicing of labial and/or velar stops, which allows us to gauge whether the model prefers a voiced or voiceless stop in the specified context.

4.3.1 Results for basic models

The log likelihood of the training data for the full, all-natural, and all-unnatural models is shown in Table 10 (closer to zero means a better fit to the data): as expected, the all-natural and all-unnatural models do worse than the full model, since they have fewer constraints at their disposal, and they do not have the mixture of natural and unnatural constraints available to represent the tendencies in the TQ data. However, the all-natural model still performs much better than the all-unnatural model. This means that the natural constraint set is generally well-suited to the TQ data set — only some unnatural elements may be needed to improve its fit.

Context	**Data points**		
Word-initial	/Pata/	→	[pata, bata]
	/Kata/	→	[kata, gata]
Post(/pre)-obstruent	/atPa/	→	[atpa, atba]
	/atKa/	→	[atka, atga]
	/aKPa/	→	[akpa, agpa, akba, agba]
Intervocalic	/aPa/	→	[apa, aba]
	/aKa/	→	[aka, aga]
Postnasal	/amPa/	→	[ampa, amba]
	/aŋKa/	→	[aŋka, aŋga]
Post-R	/alPa/	→	[alpa, alba]
	/alKa/	→	[alka, alga]

Table 9: Test dataset.

	Full	**All-natural**	**All-unnatural**
Log-likelihood	−609	−736	−1082

Table 10: Log likelihood of baseline models.

Table 11 shows how well the predictions of these three models on the test dataset[13] match the training dataset using four of the tendencies in the TQ data.[14]

As can be seen, the full model very closely tracks the proportions in the training data. The all-natural model behaves identically to the full model on the natural tendencies (word-initial devoicing and post-resonant voicing), but performs much worse on the unnatural tendencies (post-obstruent voicing and postnasal devoicing). However, unexpectedly, post-nasal devoicing is still slightly above chance (0.67) in the all-natural model (compared to 0.91 in the all-unnatural model). In Section 4.3.2, we interpret this outcome as a consequence of postnasal devoicing being parasitically motivated by the constraint *[+voice], without using the postnasal context at all.

Finally, the unnatural model behaves identically to the full model on postnasal devoicing, but performs at or near chance on the other three tendencies. As expected, it cannot capture the natural tendencies. Interestingly, though, there is no tendency for post-obstruent devoicing (as can be seen in the Supp. Materials, only a weak tendency shows up for labials, but not otherwise). This is surprising, since post-obstruent devoicing is an unnatural pattern. However, these results arise because the all-unnatural model lacks the natural CodaCond([+voice]) constraints, which are necessary to fully capture the post-obstruent voicing pattern: post-obstruent voicing involves a voiceless obstruent followed by a voiced [b] or [g]; the DisAgree constraints alone only mandate that these two differ in voicing: [tb] or [dp]; and the unnatural CodaCond([-voice]) constraints actually prefer [dp] over [tb]. Without a constraint that prefers [tb] over [dp], post-obstruent voicing cannot be properly derived.

---

[13] The predictions are calculated by averaging the probabilities of the candidates in the test dataset that conform to these tendencies.

[14] The proportions in the training data are calculated as the proportion of words that start with p or k out of all words that start with a labial or velar stop, the proportion of occurrences of a post-obstruent b or g out of all obstruent+p,k,b,g clusters, the proportion of mp and k clusters out of all nasal-p,k,b,g clusters, and the proportion of all Rb and Rg clusters out of all R+p,k,b,g clusters, respectively.

Context		**Training data**	**Full model**	**All-natural**	**All-unnatural**
Word-initial devoicing	#T	0.98	0.98	0.98	0.50
Post-obstruent voicing	VTDV	0.80	0.83	0.35	0.45
Postnasal devoicing	VNTV	0.90	0.91	0.67	0.91
Post-R voicing	VRDV	0.73	0.74	0.74	0.50

Table 11: Strength of phonotactic tendencies in basic models, as gauged by test set.

4.3.2 Results for blended models

To investigate which mix of natural and unnatural constraints is best suited to the TQ data, we examine the blended models introduced in Section 4.1.2. Table 12 shows the log-likelihood of the training data for each of the blended models. The complete results are shown in the Supplementary Materials, section A.4. Only the blended models with DisAgree (voicing disagreement) constraints, *ND (postnasal devoicing) constraints, and DisAgree & *ND constraints have a greater log-likelihood on the training data than the all-natural model. As can be seen in Table 12, the improvement in log-likelihood for the *ND model over the all-natural is much smaller than for the other two, with the other two blended models performing almost or exactly as well as the full model. In addition, the DisAgree & *ND model does not perform better than the DisAgree model, which trumps our expectation that adding *ND to the DisAgree model would improve its performance. Thus, while swapping out the unnatural *ND family of constraints for the natural *NT family constitutes a small improvement over the all-natural model, it does not constitute an improvement over the DisAgree model, which itself is the best-performing blended model.

	Full	**All-natural**	**DisAgree**	***ND**	**DisAgree & *ND**
Log-likelihood	−609	−736	−609	−714	−614

Table 12: Log likelihood of blended models.

Table 13 shows the strength of the phonotactic tendencies in the three blended models with the full model as a baseline. The numbers that deviate from the full model's results by more than 5 percentage points are shaded in grey. Interestingly, the model with only DisAgree as its unnatural constraint produces results that are practically identical to the full model. The *ND model fails to capture the prevalence of voicing-disagreeing clusters. Finally, the DisAgree & *ND model underestimates the tendency towards [kb] clusters (which is reflected in the slightly lower post-obstruent voicing score), while it overestimates the tendency towards nasal+[k] clusters (which is reflected in the slightly higher postnasal devoicing score).

Context	**Training data**	**Full model**	**DisAgree**	***ND**	**DisAgree & *ND**
Word-initial devoicing	0.98	0.98	0.98	0.98	0.98
Post-obstruent voicing	0.80	0.83	0.83	0.41	0.83
Postnasal devoicing	0.90	0.91	0.91	0.91	0.95
Post-R voicing	0.73	0.74	0.74	0.74	0.74

Table 13: Strength of phonotactic tendencies in selected blended models, with full model as comparison, as gauged by the test set.

The behavior of the *ND and DisAgree & *ND models can be interpreted as follows. The *ND model does not have access to the DisAgree family of constraints, which prefers voicing mismatch in clusters. The preference for voicing in clusters and the preference for devoicing elsewhere can no longer be differentiated by context, which leads to zero weight for *D. The postnasal devoicing pattern is captured by high-weighted *Nb and *g rather than by *D. Because there is no constraint to prefer post-obstruent voicing and there are some active constraints that prefer devoicing across the board, the model has a slight preference for post-obstruent devoicing on average.

The *ND constraint family bears zero weight in the DisAgree & *ND model, which means that *D constraint family accounts for postnasal devoicing. The *D constraint family has similar weights in the the DisAgree & *ND model and in the DisAgree model, but the DisAgree model puts some weight on the *NT (postnasal voicing) constraint family, while the *NT constraint family is absent from the DisAgree & *ND model. This is the likely cause of the "overshoot" in postnasal devoicing found in the DisAgree & *ND model.

4.4 Accounting for lack of within-word variation: indexed constraints

4.4.1 Indexed constraints setup

One important limitation of the models shown above is that they do not predict that the actual lexical items of TQ have categorical voicing, but rather, predict that most labial and velar stops will vacillate between [p/b] and [k/g], respectively. To better model this aspect of the data, we present a model with indexed constraints, which allow the simultaneous expression of lexicon-wide trends (e.g., probabilistic postnasal devoicing) and word-specific ones (e.g., the fact that t͡ʃimba- 'to put a bag on someone's back [colocar un bulto en la espalda de una persona]' is only recorded with [b] and t͡ʃimpa 'opposite side (of a watercourse) [orilla contraria]' is only recorded with [p]). This analysis could also be made in Cophonology Theory (see Sande et al. 2020 for a recent implementation of Cophonology Theory in a probabilistic framework): words with [b] and [g] go through different cophonologies than words with [p] and [k] do, but novel words are sent through these cophonologies at different rates. In this case, indexed constraints were chosen for the pragmatic reason that they can represent trend-following words and exceptions in the same grammar, which makes for simpler learning implementation and model evaluation. However, in future work, it would be important to verify whether similar results can also be reached in a Cophonology Theory analysis.

Our simulation with indexed constraints is based on the full constraint set: for each constraint in this set, an indexed version was also made. The indexed version of each constraint has violations for exactly those words (inputs) for which this constraint is winner-favoring (i.e., the winner has the fewest violations and there are losing candidates with more violations; Prince 2002); see Becker (2009) and Round (2017), among others, for earlier application of this idea. Because the indexed version of the constraint only has winner-favoring marks, it can be ranked higher than its unindexed (lexicon-wide) version, which includes both winner-favoring and loser-favoring marks. The higher-weighted, indexed version of the constraint represents those words that are exceptional undergoers of the relevant patterns, whereas the lower-weighted, unindexed version represents the role the constraint plays for other words, including novel/wug words, which have no lexical markings for these constraints. (Cf. the discussion of exceptional (non-)undergoers in Round 2017. Since all constraints in our models are markedness constraints, indexed constraints refer to exceptional undergoers of a process, but an indexed faithfulness constraint would refer to exceptional non-undergoers of a process.)

For instance, the constraint *NT is winner-favoring for all words that contain postnasal stops but no postnasal voiceless stops (e.g. t͡ʃimba- 'to put a bag on someone's back [colocar un bulto en la espalda de una persona]', wambra 'girl' [niña]). Even though this constraint is not obeyed by the entire lexicon (see the description in Section 3.3.1), those few words that do obey *NT are indexed to *NT$_i$, and a high weighting of this constraint can ensure that these particular words come out near-categorically with voiced stops after nasals.

The 42 original unindexed constraints and the 42 corresponding indexed versions, made according to the procedure described in the previous paragraph, are combined together, yielding a set of tableaux with 84 constraints for the training and test data sets.

4.4.2 Indexed constraints results

The model described above is trained and tested in the same way as the other models. This model yields a log-likelihood of -11 for the entire dataset, which is a factor of 5×10^{259} better than the model without indexed constraints (-609). This is because for all but 8 words, this model predicts all voicing values correctly with near-categorical certainty. The 8 words for which this does not happen are words in which intervocalic p and b or k and g co-occur in the same word, e.g., t͡ʃagataku- ‘to cross a road or place [atravesar, cruzar un camino o paraje]’ — in these words, voicing cannot be decided based on constraints and word identity alone. To account for these cases, something like constraints indexed to particular segments in a word would be necessary (Temkin-Martínez 2010; Round 2017).

In the test dataset, which is identical to the one for the previous simulations, and in which we assumed that none of the words are indexed to indexed constraints, we see an interesting pattern. As shown in 14, both non-indexed models perform comparably on the natural patterns of word-initial devoicing and post-obstruent voicing, but they are very different in their predictions regarding the unnatural patterns of post-obstruent voicing and postnasal devoicing, matching the pattern in the data quite well.[15] The model with indexed constraints predicts categorical post-obstruent voicing in novel words, but no postnasal devoicing tendency (in fact, it actually predicts a tendency towards postnasal voicing), whereas the model without indexed constraints predicts a tendency towards both post-obstruent voicing and postnasal devoicing, which is more in line with the data distribution. From the blended model results, we have already found that post-obstruent voicing constraints seem to be crucial for the TQ dataset. However, postnasal devoicing constraints do not seem to be as vital. This is borne out by the success of the DisAgree model and the relative lack of success of the *ND models (see Section 4.1.2). The indexed constraint analysis seems to corroborate this: it treats post-obstruent voicing as the default, and postnasal voicing as an exception to the default pattern.

Context	**Training data**	**Full model, no indexed constr.**	**Full model, with indexed constr.**
Word-initial devoicing	0.98	0.98	0.96
Post-obstruent voicing	0.80	0.83	1.00
Postnasal devoicing	0.90	0.91	0.41
Post-R voicing	0.73	0.74	0.57

Table 14: Strength of phonotactic tendencies in indexed model (with full model without indexed constraints as comparison), as gauged by the test set.

One caveat is that this simulation is run with a very weak regularization term (σ=1,000,000), which means that the model tolerates the spreading of weight among many different constraints.

---

[15] The indexed constraint model can be made to generalize in a rather more frequency-matching fashion by using Becker (2009) method of probabilistically assigning indexed constraint violations to novel forms.

As the regularization gets stronger (lower σ), weight is prioritized towards the most essential constraints, and it is predicted that the indexed (lexically specific) constraints win out over the unindexed (lexicon-wide) constraints. For novel words (in which only the lexicon-wide constraints operate), this means that the grammar will predict a distribution much closer to chance for all phonotactic tendencies, and there will not necessarily be "overlearning" for post-obstruent voicing. However, based on the reasoning above, we predict that postnasal devoicing will never be represented as stronger than post-obstruent voicing, even if regularization is stronger. To test this, we ran the same simulations with three increasingly more stringent regularization terms: σ=10,000, σ=1,000, and σ=600, with results as in Table 15: All phonotactic tendencies except post-R are "flattened towards 0.5 as the variance goes down, as predicted. Interestingly, post-R voicing is actually represented more strongly in the grammar with lower variance values, but this might simply be attributable to a lower weight of *D. In any case, it can be seen that, even with the strictest possible regularization (σ=600), post-obstruent voicing is still more strongly represented in the grammar than postnasal devoicing. This shows that when the grammar is faced with a lexicon vs. grammar choice, it will more strongly represent post-obstruent voicing in the grammar compared to postnasal devoicing.

Context	**Training data**	$\sigma = 1,000,000$	$\sigma = 10,000$	$\sigma = 1,000$	$\sigma = 600$
Word-initial devoicing	0.98	0.96	0.91	0.87	0.86
Post-obstruent voicing	0.80	1.00	0.96	0.88	0.85
Postnasal devoicing	0.90	0.41	0.48	0.50	0.49
Post-R voicing	0.73	0.57	0.60	0.64	0.67

Table 15: Strength of phonotactic tendencies in full model without indexed constraints with different regularization terms, as gauged by the test dataset.

These results demonstrate a few interesting points. Firstly, they show the feasibility and learnability of a model that combines the representation of gradient voicing tendencies (across words) along with lexically fixed voicing values (for individual words). Secondly, they show that allowing the model to take both phonotactics (natural and unnatural) and lexical factors into account predicts that the model will explicitly represent post-obstruent voicing in the grammar, but that this will not necessarily happen for postnasal devoicing. This can be attributed to words with postnasal voiceless stops getting a "free ride" on the lexically specific (indexed) version of *[+voice], so that there is no clear necessity for a high ranking of the unnatural constraints that promote postnasal devoicing.

4.5 General discussion of grammar simulations

The Maximum Entropy simulations shown in this section are aimed to test to what extent, if any, the Tarma Quechua data require unnatural constraints and to interpret learning in MaxEnt models. These are all simulations with pre-defined natural and unnatural markedness constraints and no specific role for faithfulness. Based on the literature reviewed in Section 4.1, it is assumed that there are natural constraints against obstruent voicing across the board and word-initially, and in favor of obstruent voicing after nasals, non-nasal sonorants, and intervocalically. Furthermore, we assume natural constraints for voicing agreement among adjacent obstruents (AGREE(voice)) and against voicing in codas unless followed by a voiced obstruent (CODACOND([+voice]). These constraints are inverted in ways specified in Section 4.1 to yield unnatural counterparts. We systematically compare models with natural and unnatural constraints, while keeping the number of constraints the same, and we also add a model with all constraints and a model that includes all constraints together with lexical factors.

While the best-performing model without lexical factors is the one with all constraints, out of the models with a restricted constraint count the best-performing one is the one with an unnatural constraint favoring voicing disagreement in obstruent clusters and natural constraints otherwise (the DisAgree model). The model that includes both voicing disagreement and postnasal devoicing constraints performs similarly, but the postnasal devoicing constraints receive zero weight. This implies that an unnatural constraint promoting voicing disagreement is necessary, while an unnatural constraint promoting postnasal devoicing is not: postnasal devoicing can be accounted for by giving high weight to the context-free natural constraint *[+voice].

Finally, a model with built-in lexical factors (indexed constraints that refer to all words that would be accounted for by that constraint) can ensure that there is no (notable) within-word variation while still maintaining probabilistic phonotactic tendencies in the grammar. This model consistently learns a stronger lexicon-wide tendency towards post-obstruent devoicing, compared to the lexicon-wide tendency it learns towards postnasal devoicing. This can be taken as evidence that post-obstruent voicing should be represented in the grammar as such, while postnasal devoicing may rather have a more lexical/exceptional status in the language.

Thus, while TQ exhibits two statistically significant phonetically unnatural tendencies –post-obstruent voicing and postnasal devoicing (Section 3.3)– the most plausible scenario given our simulations is that only post-obstruent devoicing has a special status in the grammar. Gradient postnasal devoicing in our dataset can be modeled as an interaction between a low-weighted constraint in favor of postnasal voicing (*NT) and a high-weighted constraint against obstruent voicing (*[+voice]), since voiced obstruents are still very much a minority in the native vocabulary. If the proportion of voiced obstruents had been higher across the board in this language while the proportion of voiced obstruents after nasal had remained the same, there might have been evidence for a phonologically unnatural constraint *ND, however. This means that, against our expectations, the Tarma Quechua phonotactic data do not provide evidence for a postnasal devoicing constraint in the phonological grammar. While TQ phonotactics do not support the *ND constraint, the *morphophonological* alternations shown in Section 3.5 that are limited to a subset of morphemes might provide independent evidence for *ND. Testing the productivity of these alternations and their theoretical implications is left for future work. Additionally, there are other languages that provide strong evidence in favor of the *ND constraint: Hyman 2001; Coetzee & Pretorius 2010; Beguš 2019.

The post-obstruent voicing pattern, however, must be supported by an “unnatural” constraint in the grammar, which we term DisAgree(voice). According to the definition in Section 1.2, an unnatural pattern is one that runs counter to a universal phonetic tendency. By extension, a constraint that consistently favors such unnatural patterns, like Disagree(voice), can be analyzed as an unnatural constraint.

One uniquely unnatural aspect of the effects of DisAgree(voice) in Tarma Quechua is that it targets two adjacent obstruents. Phonetic pressures towards voicing uniformity in these conditions operate as strong phonetic tendencies across languages (see discussion in Section 1.2). While voicing dissimilation is not an unattested process (cf. Dahl’s Law in Gikuyu, Lombardi 1995 and references cited therein, and Rendaku in Japanese, e.g., Itô & Mester 2003), it typically operates non-locally, where the articulatory phonetic tendencies towards voicing agreement are less strong. Another uniquely unnatural aspect arises from the interaction of DisAgree(voice) with the (natural) CodaCond([+voice]) constraint, namely, a preference for output clusters that contain voicing in the second member (TD) rather than in the first member (DT; see Table 6 and Section 3.3.2 for why this distribution is phonetically unnatural). No other effects attributable to DisAgree(voice) are found in our data, which means that DISAGREE(voice) in TQ consistently favors phonetically unnatural patterns, meaning that it does constitute an unnatural constraint according to our definition.

Despite its clear phonetic unnaturalness, a constraint like DISAGREE may have a functional motivation. Correspondences between consonants can be conceptualized in the Agreement-by-Correspondence (ABC) framework (e.g., Rose & Walker 2004). Whereas assimilation is accounted for by consonant-to-consonant faithfulness constraints in this framework, dissimilation may be seen as arising from consonant-to-consonant anti-faithfulness constraints (see Alderete 2001 for arguments in favor of anti-faithfulness constraints). Since anti-faithfulness can be independently motivated by factors other than simply counteracting an existing phonetic tendency, one may argue that ultimately, the TQ pattern may be accounted for by a set of "grounded" constraints (Archangeli & Pulleyblank, 1994), but such a grounding is conceptual rather than phonetic.

Finally, despite the fact that it has been shown that Maximum Entropy models are not restricted to generating phonetically natural patterns when the grammar only contains natural constraints (see, e.g., Magri & Anttila 2018), the results of the current simulations show that there is still a "natural gradience bias" in these types of models: not all gradient patterns can be modeled by interactions of natural constraints, and the absence of at least certain unnatural constraints leads to problems in modeling datasets such as the TQ one. See also Hayes (2016) and O'Hara (2022) for similar results showing that natural constraints do impose some bias on possible languages even for Maximum Entropy models.

5 Conclusion

This paper argues that phonological grammar can be probabilistic and operate in phonetically unnatural directions at the same time. We present a case of unnatural gradient phonotactic restrictions in the lexicon: Tarma Quechua stop voicing. We show that the phonotactic restrictions operate in an unnatural direction (as defined in Section 1.2): in the environments presented (postnasal, post-voiceless-obstruent, and intervocalic), the articulatorily or perceptually preferred value of [±voice] is less frequent than the dispreferred value. Based on a new corpus (which includes both published and unpublished data) we argue that the unnatural phonotactic restriction is statistically significant and phonetically real. This appears to be the first report of unnatural gradient phonotactics that target segmental features in a segmental context. In Section 3.5, we provide evidence for the productivity of these phonotactic restrictions, which suggests that they are actually present in the grammar (see Coetzee & Pater 2008 and others for arguments that gradient phonotactic patterns must be present in the synchronic grammar).

We furthermore use grammar learning simulations to test whether some phonetically unnatural gradient phonotactic restrictions can be derived as a by-product of natural constraint interactions without any unnatural constraints. We argue that at least some phonetically unnatural constraints are necessary in the grammar to model the Tarma Quechua data to their full extent. Interestingly, our results (with both lexicon-wide and lexically indexed models) provide evidence for a post-obstruent devoicing constraint, but not for a postnasal devoicing constraint. This is because postnasal devoicing can be parasitically accounted for by a general devoicing constraint, given that the data show a general dispreference for voiced obstruents, while post-obstruent voicing goes against the general devoicing trend in the data.

This result is obtained using a new approach to evaluating phonological analyses by measuring goodness of fit of individual models while keeping the number of constraints constant. This approach allows us to evaluate analyses with natural and unnatural constraints or models with and without indexed constraints. The results of this approach can serve as diagnostic for which constraints are needed for a model that provides the best fit to the data. The proposed approach also helps us interpret how the MaxEnt learners are deriving the data. For example, it has helped us understand how post-nasal devoicing can be derived without unnatural constraints. The proposed approach can be used both in further discussions on natural and unnatural constraints as well as for evaluating other phonological analyses.

In future work, we hope that our evidence for the gradient unnatural phonotactic patterns presented here will be corroborated with experimental data. Because of the relative inaccessibility of the Tarma speech community, obtaining such experimental data was outside the scope of this research. However, this type of evidence would greatly enhance the strength of our claim that Con should contain at least the unnatural markedness constraints we argue for (cf. Hyman 2001; Coetzee & Pretorius 2010 for the claim that Con should contain a constraint against [+voice] after nasals). Another related direction for future work would be to test the productivity of the limited morphophonological alternations mentioned in Section 3.5 and computationally model these patterns, to see whether they might provide evidence for the activity of a postnasal devoicing constraint (*ND) in the grammar.

Furthermore, we hope this paper will inspire exploration of other cases of gradient unnatural phonotactics, particularly ones that involve a feature other than [±voice]. Finding phonotactic restrictions that gradiently favor the phonetically unnatural variant is a challenging task which has thus far not been the main focus of phonological theory. This paper draws attention to these highly informative processes, points to their theoretical implications, and provides a quantitative framework for evaluating and interpreting weighted-constraint approaches to unnatural phonotactics. We hope this work will spark further interest in phonological processes that are simultaneously probabilistic and unnatural, and we expect that more such cases will be discovered in the future.

References

Adelaar, Willem F. H. 1977. *Tarma Quechua: Grammar, texts, dictionary.* Lisse: The Peter de Ridder Press.

Adelaar, Willem F. H., and Pieter C. Muysken. 2004. *The languages of the Andes.* Cambridge: Cambridge University Press.

Albright, Adam. 2009. Feature-based generalization as a source of gradient acceptability. *Phonology* 26.9–41.

Alderete, John D. 2001. Dominance effects as transderivational anti-faithfulness. *Phonology* 18.201253.

Anderson, Stephen R. 1981. Why phonology isn't "natural". *Linguistic Inquiry* 12.493–539.

Anttila, Arto. 2008. Gradient phonotactics and the complexity hypothesis. *Natural Language & Linguistic Theory* 26.695–729. Online: https://doi.org/10.1007/s11049-008-9052-2.

Archangeli, Diana, and Douglas Pulleyblank. 1994. *Grounded phonology.* Cambridge, MA: MIT Press.

Barry, Martin, and Ryad Teifour. 1999. Temporal patterns in Syrian Arabic voicing assimilation. *14th international congress of phonetic sciences*, 24292432.

Becker, Michael. 2009. *Phonological trends in the lexicon: The role of constraints.* University of Massachusetts Amherst dissertation.

Becker, Michael; Lauren Eby Clemens; and Andrew Nevins. 2017. Generalization of French and Portuguese plural alternations and initial syllable protection. *Natural Language & Linguistic Theory* 35.299–345. Online: https://doi.org/10.1007/s11049-016-9343-y.

Becker, Michael; Nihan Ketrez; and Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87.84–125.

Becker, Michael; Andrew Nevins; and Jonathan Levine. 2012. Asymmetries in generalizing alternations to and from initial syllables. *Language* 88.231–268. Online: http://www.jstor.org/stable/23251831.

Beguš, Gašper. 2018. *Unnatural phonology: A synchrony-diachrony interface approach.* Harvard University dissertation. Online: http://nrs.harvard.edu/urn-3:HUL.InstRepos:40050094.

Beguš, Gašper. 2019. Post-nasal devoicing and the blurring process. *Journal of Linguistics* 55.689–753.

Beguš, Gašper. 2020. Estimating historical probabilities of natural and unnatural processes. *Phonology* 37.515549. Online: https://www.cambridge.org/core/journals/phonology/article/estimating-historical-probabilities-of-natural-and-unnatural-processes/CAC83CE585B82836CBC0D3A4BCDF17EB.

Beguš, Gašper. 2022. Distinguishing cognitive from historical influences in phonology. *Language* 98.1–34. Online: https://muse.jhu.edu/article/849525.

Beguš, Gašper, and Aleksei Nazarov. 2018. Gradient trends against phonetic naturalness: The case of Tarma Quechua. *Proceedings of the 48th Meeting of the North East Linguistic Society*, ed. by Sherry Hucklebridge and Max Nelson. Amherst, MA: GLSA.

Berkley, Deborah. 2000. *Gradient Obligatory Contour Principle effects.* Northwestern University dissertation.

Blevins, Juliette. 2004. *Evolutionary phonology.* Cambridge: Cambridge University Press.

Blevins, Juliette. 2008. Natural and unnatural sound patterns: A pocket field guide. *Naturalness and iconicity in language*, ed. by Klaas Willems and Ludovic De Cuypere, 121–148. Amsterdam: Benjamins.

Blevins, Juliette; Ander Egurtzegi; and Jan Ullrich. 2020. Final obstruent voicing in Lakota: Phonetic evidence and phonological implications. *Language* 96.294–337.

Bolker, Ben. 2019. GLMM FAQ. Retrieved on June 22, 2022. Online: bbolker.github.io/mixedmodels-misc/glmmFAQ.html.

Brehm, Laurel, and Phillip M. Alday. 2022. Contrast coding choices in a decade of mixed models. *Journal of Memory and Language* 125.104334. Online: https://www.sciencedirect.com/science/article/pii/S0749596X22000213.

Buckley, Eugene. 2000. On the naturalness of unnatural rules. *Proceedings from the Second Workshop on American Indigenous Languages. UCSB Working Papers in Linguistics* 9.1–14.

Byrd, Richard H.; Peihuang Lu; Jorge Nocedal; and Ciyou Zhu. 1995. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific and Statistical Computing* 16.1190–1208.

Carpenter, Angela C. 2010. A naturalness bias in learning stress. *Phonology* 27.345–392.

Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.

Coetzee, Andries W., and Joe Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language & Linguistic Theory* 26.289–337. Online: http://www.jstor.org/stable/20533006.

Coetzee, Andries W., and Joe Pater. 2011. The place of variation in phonological theory. *The handbook of phonological theory*, ed. by Jason Riggle John Goldsmith and Alan C. L. Yu., 401–434. Blackwell.

Coetzee, Andries W., and Rigardt Pretorius. 2010. Phonetically grounded phonology and sound change: The case of Tswana labial plosives. *Journal of Phonetics* 38.404–421.

Creider, Chet A. 1968. *A morphological sketch of Tarma Quechua.* Master's thesis, University of Minnesota.

Davidson, Lisa. 2016. Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics* 54.35–50.

Docherty, Gerard J. 1992. *The timing of voicing in British English obstruents.* Berlin: Foris Publications.

Donegan, Patricia J., and David Stampe. 1979. The study of natural phonology. *Current approaches to phonological theory*, ed. by Daniel Dinnsen, 126–173. Bloomington: Indiana University Press.

Downing, Laura J., and Silke Hamann. 2021. Why phonetically-motivated constraints do not lead to phonetic determinism: The relevance of aspiration in cueing nc sequences in tumbuka. *Phonological Data and Analysis* 3.139. Online: https://phondata.org/index.php/pda/article/view/49.

Frisch, Stefan A.; Janet B. Pierrehumbert; and Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language & Linguistic Theory* 22.179–228. Online: https://doi.org/10.1023/B:NALA.0000005557.78535.3c.

Gallagher, Gillian. 2010. Perceptual distinctness and long-distance laryngeal restrictions. *Phonology* 27.435–480.

Garcia, Guilherme D. 2017. Implications of an exemplar-theoretic model of phoneme genesis: A velar palatalization case study. *Phonology* 34.4179.

Garrett, Andrew, and Keith Johnson. 2013. Phonetic bias in sound change. *Origins of sound change: Approaches to phonologization*, ed. by Alan C. L. Yu, 51–97. Oxford: Oxford University Press.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a Maximum Entropy model. *Proceedings of the workshop on variation within Optimality Theory*, ed. by Jennifer Spenader, Anders Eriksson, and Östen Dahl, 111–20. Stockholm: Stockholm University.

Gouskova, Maria; Elizabeth Zsiga; and One Tlale Boyer. 2011. Grounded constraints and the consonants of setswana. *Lingua* 121.2120–2152. Online: https://www.sciencedirect.com/science/article/pii/S0024384111001732.

Haggard, Mark. 1978. The devoicing of voiced fricatives. *Journal of Phonetics* 6.95102.

Hale, Mark, and Charles Reiss. 2008. *The phonological enterprise.* Oxford: Oxford University Press.

Hastie, Trevor; Robert Tibshirani; and Jerome Friedman. 2009. *The elements of statistical learning: Data mining, inference, and prediction.* 2 edn. New York, NY: Springer.

Hayes, Bruce. 1999. Phonetically-driven phonology: The role of Optimality Theory and inductive grounding. *Functionalism and formalism in linguistics, volume I: General papers*, ed. by Michael Darnell and Edith Moravscik, 243–285. Amsterdam: John Benjamins.

Hayes, Bruce. 2009. *Introductory phonology.* Malden, MA: Wiley-Blackwell.

Hayes, Bruce. 2016. Varieties of Noisy Harmonic Grammar. Talk given at the 2016 Annual Meeting on Phonology, University of Southern California, Los Angeles, CA, October 21-23, 2016.

Hayes, Bruce; Péter Siptár; Kie Zuraw; and Zsuzsa Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85.822–863.

Hayes, Bruce, and Donca Steriade. 2004. Introduction: The phonetic bases of phonological markedness. *Phonetically based phonology*, ed. by Bruce Hayes, Robert Kirchner, and DoncaEditors Steriade, 133. Cambridge University Press.

Hayes, Bruce, and Tanya Stivers. 2000. Postnasal voicing. Ms., University of California, Los Angeles. http://linguistics.ucla.edu/people/hayes/Phonet/NCPhonet.pdf (accessed 9 May 2018).

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. Online: https://doi.org/10.1162/ling.2008.39.3.379.

Hellberg, Staffan. 1978. Unnatural phonology. *Journal of Linguistics* 14.157–177.

Hyman, Larry M. 2001. The limits of phonetic determinism in phonology: *NC revisited. *The role of speech perception in phonology*, ed. by Elizabeth Hume and Keith Johnson, 141–186. San Diego, CA: Academic Press.

Itô, Junko, and Armin Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17.4973.

Itô, Junko, and Armin Mester. 1994. Reflections on CodaCond and alignment. *Phonology at santa cruz 3*, ed. by Jason Merchant, Jaye Padgett, and Rachel Walker, 27–46. Santa Cruz: University of California, Santa Cruz.

Itô, Junko, and Armin Mester. 2002. *One phonology or many? Issues in stratal faithfulness theory.* Tokyo: Kaitakusha.

Itô, Junko, and Armin Mester. 2003. *Japanese morphophonemics: Markedness and word structure.* Cambridge, MA: MIT Press.

Iverson, Gregory K., and Joseph C. Salmons. 2011. Final devoicing and final laryngeal neutralization. *The Blackwell companion to phonology, volume II: Suprasegmental and prosodic phonology*, ed. by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, 1622–1643. Malden, MA: Wiley-Blackwell.

Jansen, Wouter. 2004. *Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch.* University of Groningen dissertation.

de Jong, Kenneth. 2004. Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics* 32.493–516.

Kaplan, Abby. 2010. *Phonology shaped by phonetics: The case of intervocalic lenition.* University of California, Santa Cruz dissertation.

Kent, Raymond D., and Kenneth L. Moll. 1969. Vocal tract characteristics of the stop consonants. *Journal of the Acoustical Society of America* 46.1549–1555.

Kiparsky, Paul. 2006. Amphichronic program vs. Evolutionary Phonology. *Theoretical Linguistics* 32.217–236.

Kiparsky, Paul. 2008. Universals constrain change, change results in typological generalizations. *Linguistic universals and language change*, ed. by Jeff Good, 23–53. Oxford: Oxford University Press.

Kümmel, Martin. 2007. *Konsonantenwandel.* Wiesbaden: Reichert.

de Lacy, Paul. 2002. *The formal expression of markedness.* University of Massachusetts, Amherst dissertation.

de Lacy, Paul. 2006. *Markedness: Reduction and preservation in phonology.* Cambridge Studies in Linguistics. Cambridge University Press.

de Lacy, Paul, and John Kingston. 2013. Synchronic explanation. *Natural Language and Linguistic Theory* 31.287–355.

Locke, John. 1983. *Phonological acquisition and change.* New York: Academic Press.

Lombardi, Linda. 1995. Dahl's law and privative [voice]. *Linguistic Inquiry* 26.365–372.

Lombardi, Linda. 2001. Why place and voice are different: Constraint-specific alternations in Optimality Theory. *Segmental phonology in Optimality Theory: Constraints and representations*, ed. by LindaEditor Lombardi, 1345. Cambridge University Press.

Magri, Giorgio, and Arto Anttila. 2018. Does MaxEnt overgenerate? Implicational universals in Maximum Entropy Grammar. *Proceedings of the 2017 Annual Meeting on Phonology*, ed. by Gillian Gallagher, Maria Gouskova, and Sora Heng Yin. Washington, DC: Linguistic Society of America.

Martin, Andrew T. 2007. *The evolving lexicon.* University of California, Los Angeles dissertation.

McCarthy, John J. 2007. Slouching toward optimality: Coda reduction in OT-CC. *Phonological studies 10*, ed. by Phonological Society of Japan, 89–104. Tokyo: Kaitakusha.

McCarthy, John J. 2008b. The gradual path to cluster simplification. *Phonology* 25.271–319.

McCarthy, John J. 2011. Autosegmental spreading in Optimality Theory. *Tones and features: Phonetic and phonological perspectives*, ed. by John Goldsmith, Elizabeth Hume, and Leo Wetzels, 195–222. Berlin: Mouton de Gruyter.

Möbius, Bernd. 2004. Corpus-based investigations on the phonetics of consonant voicing. *Folia Linguistica* 38.5–26.

Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25.83–127.

Morley, Rebecca L. 2014. Implications of an exemplar-theoretic model of phoneme genesis: A velar palatalization case study. *Language and Speech* 57.3–41.

Myers, Scott. 2010. Regressive voicing assimilation: Production and perception studies. *Journal of the International Phonetic Association* 40.163–179. Online: http://www.jstor.org/stable/44527013.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. *The production of speech*, ed. by Peter F. MacNeilage, 189–216. New York: Springer.

Ohala, John J. 1993. The phonetics of sound change. *Historical linguistics: Problems and perspectives*, ed. by Charles Jones, 237–278. London: Longman.

Ohala, John J. 1997. Aerodynamics of phonology. *Proceedings of the 4th Seoul International Conference on Linguistics (SICOL)*, 84–91. Seoul: Linguistic Society of Korea.

Ohala, John J., and Manjari Ohala. 1993. The phonetics of nasal phonology: Theorems and data. *Nasals, nasalization, and the velum*, ed. by Marie K. Huffman and Rena A. Krakow, 225–249. San Diego, CA: Academic Press.

Ohala, John J., and Carol J. Riordan. 1979. Passive vocal tract enlargement during voiced stops. *Speech communication papers*, ed. by J. J. Wolf and D. H. Klatt, 89–92. New York: Acoustical Society of America.

O'Hara, Charlie. 2022. MaxEnt learners are biased against giving probability to harmonically bounded candidates. *Proceedings of the Society for Computation in Linguistics* 5.24.

Pater, Joe. 1999. Austronesian nasal substitution and other NC effects. *The prosody-morphology interface*, ed. by René Kager, Harry van der Hulst, and Wim Zonneveld, 310–343. Cambridge: Cambridge University Press.

Pater, Joe. 2000. Non-uniformity in English secondary stress: The role of ranked and lexically specific constraints. *Phonology* 17.237–274.

Pater, Joe. 2008. Gradient phonotactics in Harmonic Grammar and Optimality Theory. Ms., University of Massachusetts, Amherst.

Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33.999–1035.

Pater, Joe. 2010. Morpheme-specific phonology: Constraint indexation and inconsistency resolution. *Phonological argumentation: Essays on evidence and motivation*, ed. by Steve Parker, 123–154. London: Equinox.

Potts, Christopher; Joe Pater; Karen Jesney; Rajesh Bhatt; and Michael Becker. 2010. Harmonic Grammar with linear programming: From linear systems to linguistic typology. *Phonology* 27.77–117.

Prince, Alan. 2002. Arguing optimality. *University of Massachusetts Occasional Papers in Linguistics* 28.10.

Puente Baldoceda, Blas. 1977. *Fonología del quechua tarmeño*. Lima: Universidad Nacional Mayor de San Marcos, Centro de Investigación de Lingüística Aplicada.

Pycha, Anne; Pawel Nowak; Eurie Shin; and Ryan Shosted. 2003. Phonological rule-learing and its implications for a theory of vowel harmony. *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, ed. by Gina Garding and Mimu Tsujimura, 101–114. Somerville, MA: Cascadilla Press.

R Core Team. 2018. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Online: https://www.R-project.org/.

Reiss, Charles. 2018. Substance free phonology. *The Routledge handbook of phonological theory*, 1 edn., 425–452. Routledge.

Rose, Sharon, and Rachel Walker. 2004. A typology of consonant agreement by correspondence. *Language* 80.475–531.

Rothenberg, Martin. 1968. *The breath-stream dynamics of simple-released-plosive production*. Basel: S. Karger.

Round, Erich. 2017. Phonological exceptionality is localized to phonological elements: The argument from learnability and Yidiny word-final deletion. *On looking into words (and beyond): Structures, relations, analyses*, ed. by Claire Bowern, Larry Horn, and Roberta Zanuttini, 59–97. Berlin: Language Science Press.

Samuels, Bridget. 2009. *The structure of phonological theory*. Harvard University dissertation.

Sande, Hannah; Peter Jenks; and Sharon Inkelas. 2020. Cophonologies by Ph(r)ase. *Natural Language & Linguistic Theory* 38.1211–1261. Online: https://doi.org/10.1007/s11049-020-09467-x.

Smith, Caroline L. 1997. The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics* 25.471 – 500. Online: http://www.sciencedirect.com/science/article/pii/S009544709790053X.

Snoeren, Natalie; Pierre Hallé; and Juan Segui. 1992. A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics* 34.241268.

Stampe, David. 1973. *A dissertation on natural phonology.* University of Chicago dissertation.

Staubs, Robert. 2011. *Harmonic grammar in r (hgr).* Software package. Online:: http://blogs.umass.edu/hgr/.

Steriade, Donca. 1997. Phonetics in phonology: The case of laryngeal neutralization. Ms., University of California, Los Angeles.

Temkin-Martínez, Michal. 2010. *Sources of non-conformity in phonology: Variation and exceptionality in Modern Hebrew spirantization.* Boise State University dissertation.

Tesar, Bruce, and Paul Smolensky. 2000. *Learnability in Optimality Theory.* Cambridge, MA: Massachusetts Institute of Technology Press.

Vennemann, Theo. 1988. *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin.* Berlin: Mouton de Gruyter.

Westbury, John R., and Patricia A. Keating. 1986. On the naturalness of stop consonant voicing. *Journal of Linguistics* 22.145–166.

White, James C. 2013. *Bias in phonological learning: Evidence from saltation.* UCLA dissertation.

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30.945–982.

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. Online: https://doi.org/10.1162/ling_a_00285.

Wilson, Colin, and Marieke Obdeyn. 2009. Simplifying subsidiary theory: Statistical evidence from Arabic, Muna, Shona, and Wargamay. Ms., Johns Hopkins University. https://linguistics.ucla.edu/people/hayes/219/papers/WilsonObdeyn2009.pdf (accessed 22 June 2022).

Yu, Alan C. L. 2004. Explaining final obstruent voicing in Lezgian: Phonetics and history. *Language* 80.73–97.

Yu, Alan C. L. 2013. Typologizing phonetic precursors to sound change. *Language typology and historical contingency: In honor of Johanna Nichols*, ed. by Balthasar Bickel, Lenore A. Grenoble, David A. Peterson, and Alan Timberlake, 395–414. John Benjamins: Amsterdam/Philadelphia.

Zoll, Cheryl. 1998. Positional asymmetries and licensing. Ms., Massachusetts Institute of Technology.

Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language & Linguistic Theory* 28.417–472. Online: https://doi.org/10.1007/s11049-010-9095-z.

A Supplementary Materials

A.1 Data collection and annotation

Each target stop's syllable position was annotated left-to-right, with the first counted syllable therefore being the rightmost. Syllable position was furthermore annotated in two ways: firstly via syllabifying each word according to normal Tarma Quechua syllable structure, and secondly via syllabifying each word in order to maximize onsets, regardless of whether these are actually licensed in the language. The latter technique of maximizing onsets was used in order to simplify the process of analyzing target CC clusters. Tarma Quechua does not typically tolerate within-syllable CC clusters, although a stop followed by a glide or liquid seems to be permitted, at least in loans. This can be seen in examples such as *kwarta* [Sp. *cuarto*] 'room' and *gra:tis* [Sp. *gratis*] 'thanks'. Because of examples such as these, consonant + liquid or consonant + glide clusters were kept when syllabifying according to the actual phonotactics of the language. Consonant + nasal clusters were not syllabified together, as evidence is marginal that these are permitted. The data also contained several glide + vowel or vowel + glide sequences; these sequences were treated as one vowel V.

Part of speech was also annotated in two ways. The first annotation simply recorded the tagging of each source text; however, each author's tagging system did not exactly align. In order to provide consistency across part-of-speech tagging, we collapsed the tagging systems of both authors into a simple four-way system of nouns, verbs, dual noun/verbs, and "other" parts of speech (e.g. adjectives, adverbs, interjections, etc). Closed-class nouns (e.g. pronouns, numbers) were not tagged as nouns for the purpose of this analysis; instead, they were tagged as "other."

Duplicate entries between the Adelaar and Baldoceda dictionaries were treated in several ways. If multiple identical-looking entries appeared together in the same dictionary entry, they were treated as unique and not consolidated together, as in the case of homophones. If a word was entered multiple times under the same entry in a single dictionary, all were considered a single entry (e.g. a single word with both verbal and adjectival senses, listed separately, would be marked as one entry). If two entries occurred in separate dictionaries, but one appeared to be a hypernym of the other, the two were consolidated into one entry (e.g. *parba* 'food offering', and *parba* 'sweet corn' ['dulce de maiz']). Lastly, entries for which one definition seemed like a plausible semantic extension of the other were still marked as separate entries, in the absence of explicit confirmation of their relationship.

Because alveolars never undergo voicing, they were omitted from the analysis; only labials and velars were kept. In addition, word-final stops and the first members of consonant clusters almost always surface as voiceless in native vocabulary, which is why they were also excluded from the analysis. For the purpose of statistical analysis, we only test consonants in prevocalic position. Morpheme boundary and word-boundary are collapsed into one category (lexical boundary, marked with #). To test significance of the observed trends in the lexicon, we include only native vocabulary in the analysis and exclude words marked as derived in order to avoid overrepresenting patterns in those lexical items that are frequently derived (N = 1,118). We also perform a significance test on vocabulary that includes loanwords and derived items to test whether observed patterns persist when the entire vocabulary is included.

A.2 Models

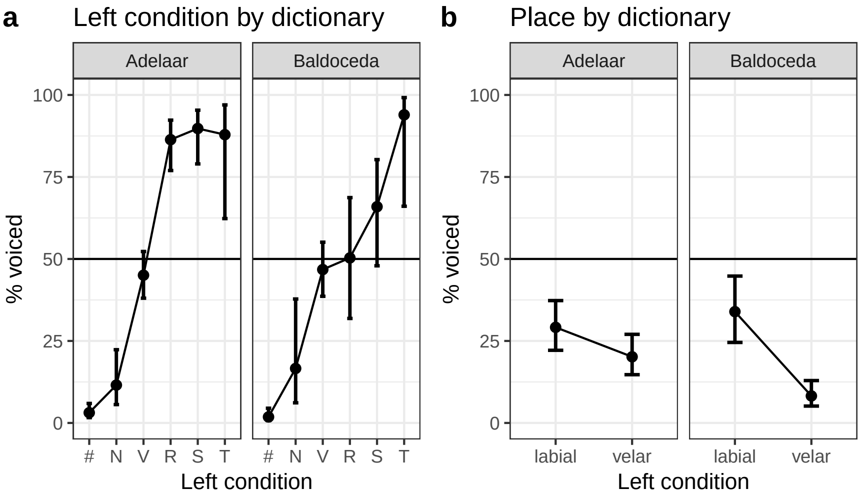


Figure 7: Estimates of a logistic regression model with Place, Left Condition, and Dictionary as predictors and two interactions: LeftCond:Dictionary and Place:Dictionary (non-loan non-derived vocabulary only).

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)=V	-0.1623	0.0852	-1.91	0.0567
#	-2.0487	0.1671	-12.26	0.0000
N	-1.7818	0.3061	-5.82	0.0000
R	1.2640	0.2147	5.89	0.0000
S	1.1647	0.2368	4.92	0.0000
T	2.3277	0.6120	3.80	0.0001
Place1	0.5264	0.0852	6.18	0.0000
#:Place1	0.2004	0.1671	1.20	0.2302
N:Place1	-0.8849	0.3061	-2.89	0.0038
R:Place1	0.2589	0.2147	1.21	0.2278
S:Place1	0.0582	0.2368	0.25	0.8059
T:Place1	-0.8460	0.6120	-1.38	0.1669

Table 16: Estimates of a logistic regression model with Place (sum-coded) and Left Condition and their interaction as predictors with loans and derived words included in the analysis.

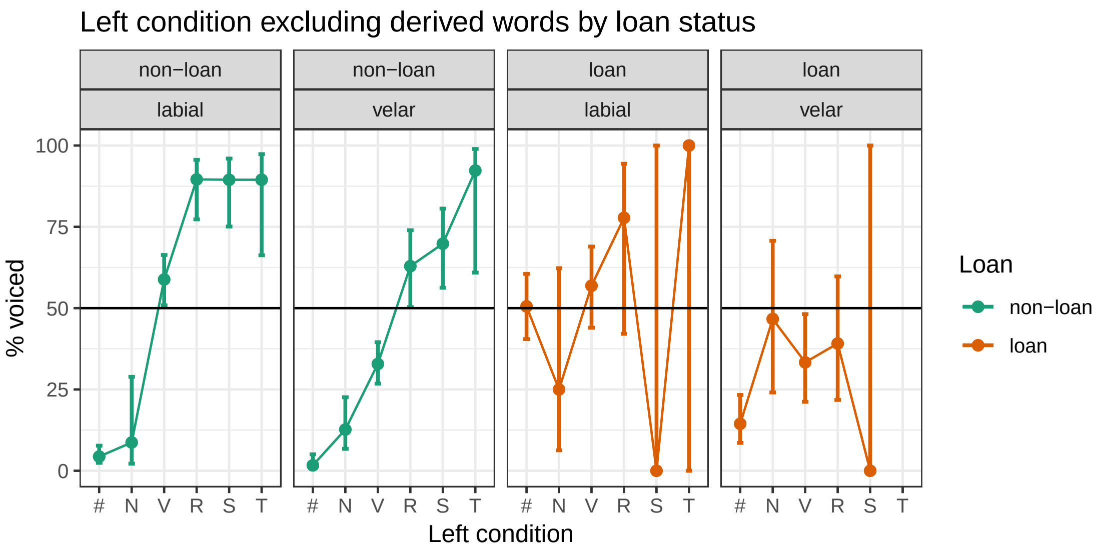


Figure 8: Estimates of a logistic regression model with three predictors, Left Condition, Place, and Loan with all interactions (derived words excluded).

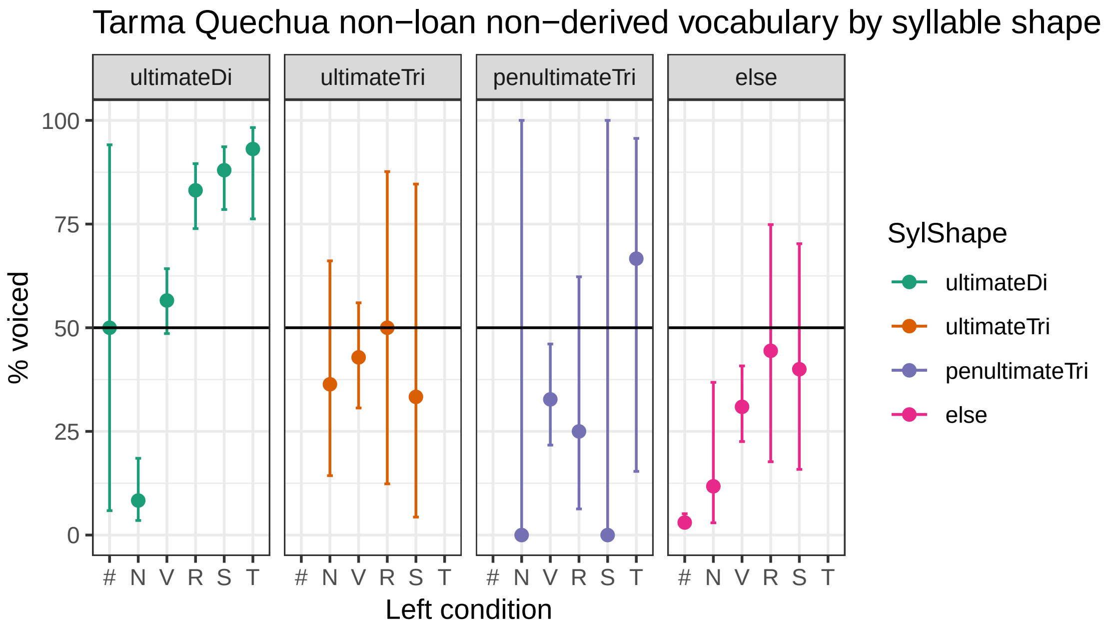


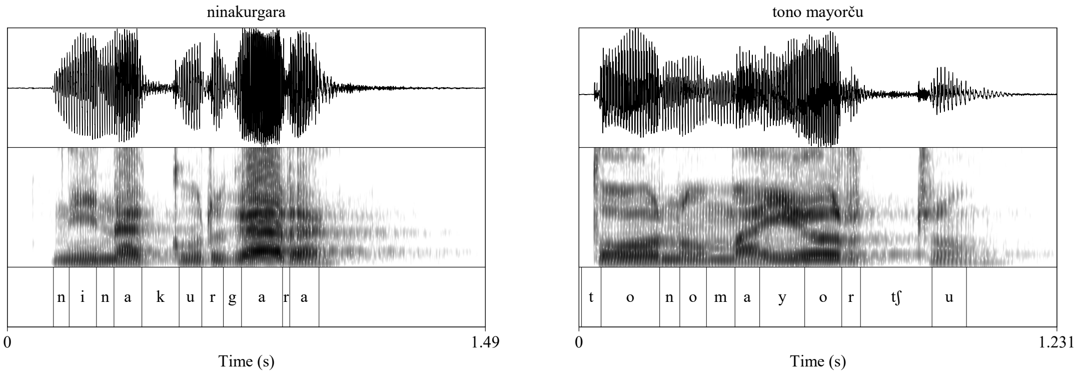
Figure 9: Estimates of a logistic regression model with Left Condition and Syllable structure as predictors with their interaction.

A.3 Analysis of reverberation

Due to generally sub-optimal recording conditions, reverberation is present throughout the audio analyzed here. However, we can still reliably distinguish between energy caused by the reverberation of the preceding vowel and voicing. This is done by comparing the suspected reverberation in two environments: word-medially and utterance-finally.

Figure 2 shows four utterances beginning with a VT sequence. The initial vowel is either [a] or [u]. We can see from these examples that following the ending of the initial vowel, energy is still visible in the waveform and spectrogram. This energy seems to last for approximately 75-100 ms, before the signal reverts to reflecting the basic background noise of the room.

If this energy were reverberation, we would expect that it would still be present utterance-finally; that is, when there is no possible following voiced sequence. If this energy were voicing, we would not expect to see it utterance-finally for the same reason. Figures 10 show the vowels from Figure 2 utterance finally. It is evident from these utterance-final vowels that some vibration is still present after the end of the vowel, again for around 75-100 ms. This supports the interpretation that this energy in both cases is reverberation, and not voicing.



(a) Utterance-final [a] in [ninakurgara].

(b) Utterance-final [u] in [mayorču].

Figure 10: Waveform and spectrogram of utterance-final [a] and [u], demonstrating reverberation.

That the energy present in the initial consonant segment of TD clusters in Figure 2 is not voicing can also be shown through comparing each putative T segment with its putative D segment. The voiced segments each have a noticeably darker bar at the bottom, low-frequency end of their respective spectrograms, indicative of voicing. Clear periodicity can also be seen on their respective waveforms throughout the segment, again indicative of voicing. While the T segments do show some periodicity in the beginning of their waveforms, this is not unexpected, and likely represents the periodic reverberation of the previous vowel during the articulatory transition to a voiceless stop.

A.4 Data log-likelihood of all non-indexed grammar models

Full	**All-natural**	**All-unnatural**	
−609	−736	−1082	
***[-voice]**	***#[-voice]**	**DisAgree**	**CodaCond([-voice])**
−794	−820	−609	−763
***VDV**	***ND**	***RD**	**DisAgree & *ND**
−742	−714	−777	−614

Table 17: Log likelihood of all non-indexed grammar models.