Examining learnability of grammar
A view from Kashmiri loanwords

Sneha Ray Barman

In this paper, I address the issue of phonological learnability with a focus on Kashmiri loanwords. The variation of Kashmiri recorded in this paper exhibits two repair strategies while borrowing words from Hindi-Urdu, English, and Persian; namely, aspiration of voiceless stops at the coda position and epenthesis of a vowel in onset consonant clusters. I have attempted to analyze these issues using Optimality Theory. Checking the learnability of the proposed grammar, I show that it is possible to learn the grammar but the ranking differs when the model is examined using the Recursive Constraint Demotion and Maximum Entropy Grammar.

1. Introduction

Computational models of language learning are used to comprehend how linguistic knowledge is present in the human brain, how it is learned, and the constraints of learning and variation. The central concept of the phonotactic model centers around sound rather than words, in contrast to conventional language models. In an effort to match gradient phonotactic knowledge, computational representations of phonotactic information are created. Tessier (2017) examined the formal properties of phonological languages and grammar in relation to algorithms that seek to learn the language-specific elements of grammar. The computer-simulated models thus try to mimic human learning ability (Albright & Hayes 2011). In this paper, I have addressed the case of Kashmiri loanwords and how phonological alternations make them stand out from other Indo-Aryan languages. Ramadoss & Vijayakrishnan (2006) thoroughly studied the role of epenthesis in Kashmiri English. The data recorded in this paper reflects /ɨ/ as the epenthetic segment, while Ramadoss and Vijakrishnan referred to /ə/ as the epenthetic vowel. Due to the lack of acoustic data, I have treated this as a dialectal difference. By doing a learnability experiment, I hope to shed light on the fact that probabilistic models like Maximum Entropy Grammar present with more authentic constraint ranking than the universal approach of Optimality Theory.

Section 2 is a brief introduction to Kashmiri phonology and its basis within the OT frame-

Proceedings of ConSOLE XXXI, 2023, 237–251
https://www.universiteitleiden.nl/en/events/series/sole

work. Section 2.1 explains the treatment of the Kashmiri syllable structure within the OT framework. Section 3 analyses the loanwords in Kashmiri with the help of OT. It is divided into two sub-sections to treat the cases of aspiration and epenthesis separately. Sections 4 and 4.1 examine the learnability of the proposed grammar using Recursive Constraint Demotion. Section 5 gives a brief background on Maximum Entropy Grammar and Section 5.2 checks the constraints using Maximum Entropy Grammar. I conclude in Section 6.

2. Kashmiri phonology: an overview

Kashmiri, locally called Koshur (/kəːšur/), is spoken across the valley of Kashmir in India and Pakistan. Kashmiri differs from other Indo-Aryan languages in certain phonological, morphological, and syntactic aspects. Grierson (1919) pointed out a few cases that make Kashmiri stand out from the Indo-Aryan crowd. Some of them are:

- Kashmiri is the only SVO language in the Indo-Aryan language family.
- It does not have a four-way contrast of plosives, that is, Kashmiri lacks voiced aspirated plosives but preserves the voiceless ones. The four-way contrast of stops is a very common feature of Indo-Aryan languages.
- Apart from these, the existence of central vowels like /ɨ, ɨː, ə, əː/ is also unique to Kashmiri (Koul 2005).
- It is a V2 language [(O)V(C)] like German, Dutch, Icelandic, and a few other languages while also showing some strikingly unique features. Another interesting fact is that all the oral vowels in Kashmiri can be nasalized. It is denoted by the nasal (~) sign over the vowels.

Vowel	**Front**	**Central**	**Back**
High	i iː	ɨ ɨː	u uː
Mid	e e:	ə əː	o oː
Low		ɑ ɑː	ɔ ɔː

Table 1. Vowel inventory of Kashmiri

Aspiration: Aspiration is phonemic in Kashmiri. The phonemic inventory contains voiceless aspirated stops only; voiced stops cannot be aspirated in the language. They can occur in all three positions in a word. The examples have been laid out in Table 3.

Consonant clusters: Consonant clusters are common in the language. Although it comes with a few restrictions, the clusters can occur in all three-word positions, i.e., initial, medial, as well as final positions. The current paper focuses on word-initial and word-final clusters and discusses them in detail.

Word-initial consonant clusters: Kashmiri allows complex onset clusters. The maximum number of segments in Kashmiri complex onset is two (CCV. . .). However, the sequence is not abrupt. The word-initial consonant clusters, although not very frequent, follow a pattern. In the

Consonants	Bilabial	Alveolar	Retroflex	Palatal	Velar	Glottal
Stops	p b	t d	t̥ d̥		k g	
aspirated stops	p^{h}	t^{h}	t̥^{h}		k^{h}	
Affricates		ts	j	č		
aspirated affricates		ts^{h}		č^{h}		
Nasals	m	n				
Fricatives		s z		š		h
Laterals		l				
Trills		r				
Semi-vowels	v	y				

Table 2. Consonant inventory of Kashmiri (Koul 2005)

Kashmiri	Gloss	Kashmiri	Gloss
[p^{h}al]	fruit	[paːp^{h}]	sin
[sap^{h}eːd]	white	[t^{h}od]	tall
[t̥^{h}ul]	egg	[zuːt̥^{h}]	tall
[mit̥^{h}əːy]	sweets	[k^{h}ɔk^{h}ur]	hollow
[mat^{h}un]	rub	[k^{h}anun]	dig
[sat^{h}]	seven	krak^{h}	cry

Table 3. Aspiration in Kashmiri (Koul 2005)

following examples, we can see that the first segment of the consonant cluster is less sonorous than the second one. Therefore, the onset cluster follows the Sonority Sequencing Principle (henceforth, SSP).

Kashmiri	Gloss	Kashmiri	Gloss
[proːn]	old	[p^{h}ras]	poplar tree
[broː]	cat	[tre]	three
[drog]	expensive	t̥roːt̥^{h}	trout(fish)
[d̥ram]	drum	[krãz]	skeleton
[k^{h}raːv]	footwear	[srog]	cheap
[ts^{h}rat̥^{h}]]	mischief	[šraːn]	brathroom

Table 4. Onset clusters in Kashmiri (Koul 2005)

The most salient feature of Kashmiri word-initial consonant clusters, as evident in the above examples, is that the second member of the cluster is always /r/, and the preceding member is a stop /p, p^{h}, b, t, t^{h}, t̥, d̥, k, k^{h}, g/, affricate /ts^{h}/ or a fricative /s, š/. Therefore, the sequence goes from less sonorous (stops/affricates/fricatives) to more sonorous (/r/), resulting in a sonority increment.

Word-final consonant clusters: Word-final clusters in Kashmiri are evident. The maximum number of segments is two (…CC). It also follows a pattern where the first member of the cluster is always a nasal (/m/, /n/) or a fricative (/s/, /š/), and the second member is always a stop /p, p^{h}, b, t, k, etc/. Therefore, the coda clusters decrease in sonority [nasal/fricative (more sonorous) $\rightarrow$ stops (less sonorous)]. This means the Kashmiri word-final consonant clusters follow the SSP.

Kashmiri	Gloss	Kashmiri	Gloss
[laemp]	lamp	[šank^h]	conch
[amb]	mango	[kašt̥]	trouble
[dand̥]	teeth	[mast]	carefree
[k^{h}and̥]	sugar	[gašt]	round

Table 5. Coda clusters in Kashmiri (Koul 2005)

2.1. Treatment of Kashmiri syllable structure in OT

The Optimality Theory grammar was developed by Prince & Smolensky (2004) and it gives language learners the responsibility of identifying a grammar that is consistent with the target language. Through the lens of constraints, this grammar examines phonological interactions. It adopts constraint-based phonology in place of the rule-based phonology used in generative grammar. It is assumed that every linguistic output form satisfies the set of ranked constraints in the best possible way or in the "most harmonic" way. According to Kager (2004), a candidate is said to be optimal if it incurs the fewest significant violations of a group of conflicting constraints. Therefore, OT evaluates a single input given a set of constraints and generates an infinite set of output candidates. The OT grammar is conceptually more universal in that it discusses grammar more frequently than phonology unique to a particular language.

Onsets and codas are optional in Kashmiri, and the native grammar allows consonant clusters in all three positions. Thus, well-formedness constraints ONSET (a syllable must have onset), NO-CODA (syllables are open) (Itô 1989; Prince & Smolensky 2004) and *COMPLEX (complex onset and coda are not allowed) are dominated.

Koul (2005) explained that the word-initial consonant clusters in Kashmiri are grammatical only if a stop/affricate/fricative is followed by /r/, while word-final clusters are acceptable only if a nasal/fricative is followed by a stop. However, the existence of loanwords like /kla:b/ 'club' or /ple:n/ 'aeroplane' suggests that the clusters follow the SSP (initial clusters should rise in sonority while the final clusters should fall). I use the constraint SONORITY as a markedness constraint. SONORITY assigns a violation to the syllable that does not follow the SSP. I use IDENT-IO (the output must have correspondent segments in the input) as the faithfulness constraint. Since a consonant cluster is ungrammatical when Sonority is violated, it is the highest-ranked constraint in the grammar followed by IDENT-IO. *Complex, NO-Coda, and Onset are already dominated owing to the reasons cited above. Violating the highest-ranked constraint discards the representation in grammar. In other words, the grammar does not allow initial clusters that do not abide by SSP.

Therefore, constraint ranking: Sonority » Faithfulness » *Complex, No-Coda, Onset (see Table 6).

In this section, I have shown how the case of consonant clusters in Kashmiri can be handled with OT constraints. Since aspiration is phonemic in the language and has no effect on the syllable structure, I have not delved into that aspect in this paper. As the paper proceeds, I have shown how aspiration plays a significant role in the loanword phonology of Kashmiri and propose the OT constraints to account for the same.

Input:/drog/	SONORITY	IDENT-IO	*COMPLEX	NO-CODA	ONSET
⇨ a.*drog*			*	*	*
b. *dro*		!*	*		*
c. *dirog*		!*		*	*

Table 6. OT analysis of general syllable structure in Kashmiri

3. Kashmiri loanwords

Just like the other Indo-Aryan languages, Kashmiri borrowed largely from Sanskrit, Hindi-Urdu, Persian, Perso-Arabic, and recently English. Koul (2005), however, doubted whether Arabic borrowings took place only through Persian or directly. In the current section, I will discuss the Persian (including Perso-Arabic), Hindi-Urdu, and English lexical borrowings in light of their phonological alternations. Several phonological and morpho-phonological changes take place during the borrowing phase. There are cases of vowel harmony, sound change, elision, and many more Koul (2005), however, I will focus mainly on:

- The aspiration of the voiceless stops at the coda position. I make a brief comment on the repair strategy adopted by other languages like Korean and propose how OT constraints can handle the factor (section 3.1).
- In section 3.2, I discuss the epenthesis of a vowel in word-initial consonant clusters with reference to Bangla, Turkish, Korean, etc. I also analyze the syllable structure alternated in the loanwords with the help of OT.

3.1. OT analysis of aspiration in Kashmiri loans

The data from Table 7 exhibits a set of clear examples of aspiration in borrowed words in Kashmiri. It can be noticed that voiceless unaspirated plosives (/k/, /p/, /t/) are aspirated (/k^{h}/, /p^{h}/, /t^{h}/) in the syllable-final or coda (/pa:k^{h}/, /pa:p^{h}/, /minat̥^{h}/ etc.) position.

Persian	Kashmiri	Gloss	Hindi-Urdu	Kashmiri	Gloss	English	Kashmiri
/paːk/	/paːk^{h}/	*pure*	/mulk/	/muluk^{h}/	*country*	doctor	/daːk^{h}t̥ar/
/caːlaːk/	/caːlaːk^{h}/	*clever*	/paːp/	/paːp^{h}/	*sin*	minute	/minat̥^{h}/
/naːzuk/	/noːzuk^{h}/	*delicate*	/raːt/	/raːt^{h}/	*night*	rate	/reːt̥^{h}/
/poːšaːk/	/poːšaːk^{h}/	*dress*	/davaːt/	/davaːt^{h}/	*inkpot*	paper	/peːpar/

Table 7. A partial list of loanwords in Kashmiri

A similar case occurs in Korean where the voiceless unaspirated stops become aspirated in English loanwords (e.g. [strɛs] > ‘stress’>[sɨthɨresɨ]). Kang (1996) attributed this instance to perceptual level matching, that is, the voiceless unaspirated stop /t/ is matched with the voiceless aspirated stop /t^{h}/ in Korean at the perceptual level. In response to this assertion, Lee (2000) argued that the realization of laryngeal features [voices, aspiration, and glottalization] of English obstruents in Korean is captured by the interaction of markedness constraints, prohibiting elements that require articulatory effort and faithfulness constraints requiring to preserve the input forms. To minimize disparities between the phonetic output of English and its corresponding loanword form in Korean, he proposed that a faithfulness constraint MAX [+long VOT] plays a significant role. This constraint captures similarities in the release importance of English and Korean stops.

I propose the constraint *T[-vce,-sg]syll that assigns violation to voiceless unaspirated plosives at the syllable-final position. Interestingly, Kashmiri allows a voiceless stop in the coda-position when it is a member of a cluster (e.g., /mast/ or /læmp/). The data presented in this paper does not exhibit a case where the voiceless stop is generally allowed in the coda position. Therefore, it is a marked situation in the language violation which renders ungrammaticality. *[+spread glottis] is a context-free markedness constraint that disallows aspiration, and IDENT-IO is a faithfulness constraint in that correspondents in input and output have identical features. Since aspiration of voiceless stops is a marked feature of Kashmiri, it is ranked as the highest constraint that violates the faithfulness constraint IDENT-IO as well as *[+spread glottis], and yet outputs a grammatical element. Constraint ranking: *T[-vce,-sg]syll » IDENT-IO »*[+spread glottis].

Input:[čaːlaːk]	*T[-vce,-sg]syll	IDENT-IO	*[+spread glottis]
a./*čaːlaːk*/	*!		
⇨b./*čaːlaːk*^{h}/		*	*

Table 8. OT analysis of aspiration in Kashmiri loanwords

3.2. OT analysis of epenthesis in Kashmiri loans

The data in Table 9 illustrates a case of epenthesis in onset and coda clusters in Kashmiri. /ɨ/ is inserted when consonant clusters violate the Sonority Sequencing Principle. Lombardi (2003) proposed that the epenthetic vowel is the least marked vowel possible given the contents of the language's vowel system. She further argued that a language will use the least marked vowel as an epenthetic vowel. When there are vowels like /ɨ/ or /ə/ in a language, it will always choose /ɨ/ as it is the least marked vowel. In the absence of /ɨ/, the language will choose /i/ as the epenthetic vowel. Korean loanwords show a preference for /ɨ/ (e.g., pat → p^{h}ætɨ; tube → t^{h}jupɨ) (Kang 2003); Tamil also prefers /ɨ/ as an epenthetic vowel after word-final liquids ([vɑlɨ] 'tail'), Bangla has chosen /i/ owing to the lack of both /ɨ/ and /ə/ in the inventory. However, Turkish shows an exception. It prefers /i/ as an epenthetic vowel despite having/ɨ/ and /ə/ in the grammar. According to a 2009 acoustic phonetic study by Gouskova and Hall, for some speakers, epenthetic [i] has a lower second formant value and is much shorter in duration than lexical [i]. Because of the low F2, which suggests that the articulation is relatively back, [ɨ] would be a more accurate transcription.

In Indo-Aryan languages like Bangla (Kar 2013; Nagarajan 2014) and Punjabi (Mahmood et al. 2011), initial clusters are not allowed. They have adopted inserting /i/ and /ə/, respectively, to avoid the clusters. Kar (2013) stressed examining the epenthesis in clusters consisting of coronal [/s/+stop] (e.g., /sku:l/– /iskul/). A repair strategy similar to Kashmiri is seen in Lenakel (/t-n-ak-ol/>/tɨ.na.gol/) (Kager 2004). The general well-formedness constraints I propose are SONORITY and DEP-IO (no epenthesis allowed) (Kager 2004).

Why not (ɨskul or skulɨ)?- Alignment constraints are introduced to avoid initial or final

English	Kashmiri	Gloss
/eɜpleɪn/	/ple:n/	aeroplane
/klɜk/	/klərɨk/	clerk
/isku:tə/	/sɨku:tar/	scooter
/steɪʃən/	/sɨt̥esa:n/	station
/spi:d/	/sɨpi:d/	speed
/sku:l/	/sɨku:l/	school
/nɜ:s/	/narəs/	nurse
/lɪpstɪk/	/lɪpsɨtɪk/	lipstick

Table 9. A partial list of epenthesis in Kashmiri

Input: /spi:d/	Dep (MF)	Align-L	Sonority	Dep-IO	Ident-IO	Align-R
a. spi:d			*!			
⇨b.sɨ.pi:d				*	*	*
c. sə.pi:d	*!			*	*	*
d. ɨs.pi:d		!*	*	*	*	

Table 10. OT analysis of epenthesis in Kashmiri loanwords

epenthesis. ALIGN-L restricts epenthesis on the left edge of the prosodic word, and ALIGN-R restricts epenthesis on the right edge.

Why /sɨku:l/ and not /səku:l/?- /ɨ/ appears to be the epenthetic vowel with features [+HIGH,-BACK,-LOW,-ROUND]. We can count these as the four features as marked in the language. According to Shademan (2002) DEP(MF), a surface instance of [+LOW], [+HIGH], [+ROUND], or [+BACK] must have an identical underlying correspondent. In the example of /spi:d/ > /sɨ.pi:d/, we can see that the input vowel segment /i/ has the features [+high,-back,-round]. According to Dep (MF), it is suggested that the surface representation must have an identical underlying correspondent. Therefore, the output vowel must have at least one feature similar to the input. /ə/ is a [+mid] vowel, immediately losing the opportunity to surface as an epenthetic segment, while /ɨ/ bears [+high, -round] features identical to the input /i/. Similarly, for /sku:l/>/sɨkul/, /ɨ/ shares [+high] feature with the input segment /u/ with features [+high, +back, +round] while /ə/ shares none. This makes /ɨ/ the most preferred epenthetic segment in Kashmiri.

Proposed constraint ranking: Dep (MF)» Align-L, Sonority » Dep-IO, Ident-IO, Align-R (see Table 10).

4. Checking learnability of the proposed grammar

Tesar & Smolensky (1995) developed a learnability algorithm named Recursive Constraint Demotion (RCD) following the OT concepts. The central questions of this learnability model were:

1. Given a set of surface forms of the target language and a set of universal constraints, is it possible for the learner to discover the correct constraint ranking?

2. What strategies do the learners use to converge into the proper ranking?

The basic idea of this learnability algorithm is that the information about constraint ranking can be extracted from the violation of constraints rather than satisfaction with the optimal candidate. The learner is fed with the input, assuming that a child learning their native language has access to positive evidence or the grammatical form only. The grammar is deduced by ranking the constraints into a hierarchy under which the input is the most harmonic output form out of all other generated forms. The algorithm is rather simple, following the steps below:

- Initially, all the constraints are unranked and all winner-loser pairs are unexplained. It is referred to as the linear stratum. C1, C2, C3, ... , Cn
- Next, each suboptimal candidate is compared against the optimal candidate (subopt < opt) to deduce the winner and loser marks. These are called Mark-data pairs.
- The learner then demotes all of the constraints that prefer losing candidates so that they are overpowered by constraints that prefer the winning candidates at each stage. In other words, $^*C_{\text{winner}}$ (constraints violated in optimal candidates) is demoted to a stratum immediately below $^*C_{\text{loser}}$ (constraints violated in suboptimal candidates).
- It then checks to see which winner/loser pairs have been successfully explained by having a winner-preferring constraint ranked above all loser-preferring constraints because a constraint that is violated in the optimal output must be dominated by some other constraint.
- Once a pair has been explained, it may be removed from consideration. This reduces the set of unexplained losers and (ideally) also reduces the set of loser-preferring constraints, freeing up some constraints for ranking in the subsequent stage. This demotion is recursive, that is, it is repeated until no further demotions occur.

Tesar & Smolensky (1995) prove that this learning algorithm from a single input can converge to a state in which all information from this output has been put to maximal use. However, the demotion must be minimal because maximal demotion is unable to converge into the target grammar due to its ever-changing nature.

4.1. Implementing recursive constraint demotion

Assumptions: A child only has access to the optimal candidates (positive evidence) of the grammar in their linguistic environment.

- Constraint reranking strictly adheres to the violation of constraints instead of satisfaction. A top-ranked constraint can be violated by an optimal candidate as long as it is violated at least as many times as it is violated by other outputs.
- The algorithm demotes constraints in winner-marks immediately below the constraints in loser-marks. The demotion must be minimal, hence reverse ranking is suggested (Kager 1999).

	Subopt	<	Opt
a<b	spi:d	<	sɨ.pi:d
c<b	sə.pi:d	<	sɨ.pi:d
d<b	ɨs.pi:d	<	sɨ.pi:d

Table 11. Comparison table of subopt<opt

Step 1: Linear stratum (for candidates in Table 11): All the constraints are unranked. It is called the mother stratum or the linear stratum.

H_0: {AlignL, AlignR, Contiguity, Dep-IO, Dep (MF), Sonority}

Step 2: Comparing subopt<opt: In the OT analysis shown in Table 10, the output candidate [sɨ.pi:d] is optimal, while the other ungrammatical candidates are called suboptimal with respect to [sɨ.pi:d]. A table comparing the suboptimal and optimal candidates is given in Table 11.

Step 3: Mark-data pairs table: Each suboptimal candidate is compared (based on harmony < opt) against the optimal candidate to deduce the loser and winner marks. For each pair in Table 11, the learning algorithm first builds an overview of the constraints that have been violated. The outcome is summarized in Table 12.

The marks for one pair in (11), are represented by each row. Each candidate pair has two cells, one in the column winner-marks that lists all violations for the ideal candidate, and one in the column loser-marks that shows all violations for the suboptimal candidate. Before the name of the violated constraint, an asterisk is used to denote violations. Candidates who cause several violations of one restriction are given a second mark for that constraint.

Mark-data pairs(subopt<opt)	Loser-marks	Winner-marks
a<b:spi:d<sɨ.pi:d	*Sonority	*Ident-IO,*Dep-IO,*Align-R
c<b:sə.pi:d<sɨ.pi:d	*Dep(MF),*Ident-IO,*Dep-IO,*Align-R	*Ident-IO,*Dep-IO,*Align-R
d<b:ɨs.pi:d<sɨ.pi:d	*Sonority,*Align-L,*Ident-IO,*Dep-IO	*Ident-IO,*Dep-IO,*Align-R

Table 12. Mark-data pairs

Step 4: Mark-data cancellation:Prior to "purifying" the information, the algorithm must first eliminate marks from the table that have no information value. The first thing to do is remove any scores that the winner and loser have in common from this table. Shared violations can never produce harmony discrepancies between two candidates; hence, they cannot reveal information regarding constraint ranking according to OT logic. The elimination of violation marks that have no informational value, as we already indicated, "purifies" the raw data in Table 12. Before we dive into how the algorithm accomplishes this task, let's have a look at the updated mark-data pairs in Table 13, where canceled marks have been removed.

Mark-data pairs(subopt<opt)	Loser-marks	Winner-marks
a<b:spi:d<sɨ.pi:d	*Sonority	*Ident-IO,*Align-R,*Dep-IO
c<b:sə.pi:d<sɨ.pi:d	*Dep(MF), *Ident-IO *Align-R,*Dep-IO	*Ident-IO,*Align-R,*Dep-IO
d<b:ɨs.pi:d<sɨ.pi:d	*Sonority, *Align-L,*Ident-IO,*Dep-IO	*Ident-IO,*Align-R,*Dep-IO

Table 13. Mark-data pair cancellation

Mark-data pairs (subopt < opt)	Loser-marks	Winner-marks
a< b : spi:d <sɨ.pi:d	*Sonority	*Ident-IO, *Align-R, *Dep-IO
c<b: sə.pi:d<sɨ.pi:d	*Dep (MF)	–
d < b: ɨs.pi:d<sɨ.pi:d	*Sonority, *Align-L	*Align-R

Table 14. Mark data-pair after cancellation

Step 5: Constraint Demotion: H_0 is the initial strata where the constraints remain unranked.

{Align-L, Align-R, Contiguity, Dep-IO, Dep (MF), Sonority}

In the following steps, the constraints in the winner-marks are demoted immediately below the loser marks denoting domination of higher-ranked constraints over others.

• At H_1, the suboptimal candidate (a) [spi:d] is less harmonious than the optimal candidate (b) [sɨ.pi:d]. The suboptimal candidate violates the constraint *Sonority, while the optimal candidate violates *Ident-IO, Align-R, *Dep-IO, *Contiguity. Following the rules of constraint demotion, the winner-marks constraints are demoted immediately below the H_0 stratum.

Align-L, Dep (MF), Sonority
»
Align-R, Ident-IO, Dep-IO

In the first step itself, the algorithm has already specified two distinct strata of dominated and undominated constraints. Since the target grammar is yet to be reached, the learner keeps going back to the first pair of Table 15 and rewinds the process and it continues till the target grammar is achieved. This is why this algorithm is called Recursive Constraint Demotion.

• At H_2, candidate (c) [sə.pi:d] is compared against the optimal candidate (c) [sɨpi:d]. The highest loser-mark is *Dep (MF) but there is no loser-mark. Hence, it remains the highest dominating constraint. This pair is not taken into account in this phase.

• At H_3, candidate (d) [ɨs.pi:d] is compared against the optimal candidate. At this stage, *Sonority and *Align-L are the highest-ranking constraint dominating *Align-R. Since *Align-R was already demoted in H_1, no further demotion is required.

• At H_4, the learner again compares the first pair of mark-data pairs a<c. *Sonority is the highest-ranked loser-marks constraint in the hierarchy, while *Ident-IO, *Dep-IO, and *Align-R are winner-marks. They are all dominated by *Sonority, therefore, no changes are made. So, $H_4 = H_1$.

Align-L, Align-R, Dep (MF), Sonority
»
Ident-IO, Dep-IO, Align-R

• At H_5, b<c is assessed. Dep (MF) is still the undominated loser-mark. The algorithm demotes the other two undominated constraints immediately below Dep (MF) so that the target grammar can be achieved. Therefore, a new stratum is created with this demotion.

Dep (MF)
»
Align-L, Sonority
»

Ident-IO, Dep-IO, Align-R

• At H6, d<c is assessed again. The highest constraints are the loser-marks, dominating the winner-mark *Align-R which is already at the bottom stratum. Therefore, no further demotion is performed. The learner has already achieved the target grammar. No recursive demotion is performed anymore.

Step 6: Ranking the constraints : According to RCD, the constraints should be ranked as: Dep(MF) » Align-L, Sonority » Align-R, Dep-IO, Ident-IO.

Therefore, the algorithm appears to converge into the grammar posited by OT. However, Tesar & Smolensky (1995) algorithm is proved to converge even if any one of the constraints leads to the actual ranking of the dataset, therefore, I examine them using MaxEnt grammar.

5. Maximum Entropy Grammar
5.1. Background

Though RCD appeared to converge into a target grammar, it had some limitations that led researchers to develop stochastic models of OT. The key assumption of RCD that the learner is presented with only the positive evidence of a language is an oversimplification (Kager 2004). The expectation of input grammar to be free of errors or variation makes it too unreal to be learned in a real-life situation (Albright & Hayes 2011). The claim that the total rank of constraints is available to the child is understandably controversial as well. Moreover, this OT model fails to work with real-world data that can be noisy and can contain free variations. To overcome these issues and the restrictiveness of OT, Boersma (1997) proposed Gradual Learning Algorithm (GLA), a stochastic model, which could learn from noisy training data and could generate free variation in the grammar. Discarding a set of discrete rankings, the GLA assumes a continuous scale of constraint strictness (Boersma & Hayes 2001). Constraints are arranged in ranking values on a numeric scale. The ranking values define the means of Gaussian probability distributions, from which sampling takes place when grammar is applied (Hayes 2007). However, GLA was unable to account for the effects of cumulative constraint interactions. Maximum Entropy grammar, proposed by Goldwater and Johnson in 2003, is a probabilistic framework that does away with OT notions. MaxEnt grammar is driven to use as much data from the training set as possible without assuming any additional information. In contrast to OT, it uses a weighted approach rather than ranking the constraints. Each constraint carries a weight, which is a non-zero real integer. A constraint's weight indicates the probability reduction for the candidate who deviates from it. The model has two steps that incorporate math:

1. finding the best weight for the constraints from the training data
2. observing what the resulting model is predicting for both the training data and the potential new testing data

Every candidate x has a score or harmony value (h), where that score is the sum of weighted constraint violations:

$$h(x) = \sum_{i=1}^{N} \omega_{\mathrm{i}} C_{\mathrm{i}}(x)$$

Here, h(x) is the overall harmony value of a candidate x which is determined by the summation of the numerical weights multiplied by the constraint violations incurred by each candidate. The idea is that higher-ranked constraints have greater weights than any lower-ranked constraints. The harmony values are then converted into MaxEnt score where the highest MaxEnt score incurs the lowest harmony value, thereby leading it to the optimal candidate. The maxEnt score is a negative exponential of harmony value:

$$P \cdot (x) = e^{-(h(x))}$$

MaxEnt harmony defines conditional probability P(xly) of an output y given an input x (Jarosz 2019).

$$P_{\mathsf{xly}} = \frac{P * x}{z}$$

, where Z is a normalizing constant to ensure the conditional probabilities sum to 1 for each input.

$$Z = \sum_{y \in \Omega} P * (y)$$

, where Ω is every possible form of a candidate (Hayes & Wilson 2008;Jarosz 2019).

Because of its gradience accountability and capacity for handling noisy data, the MaxEnt grammar gets preference over all other stochastic versions of OT. The learner can converge into the target grammar thanks to its weight-assigning feature. The grammar does not assume the target grammar has any constraint ranking because it evaluates potential words based on the weighted total of the violations of the constraints. Following the tenets of MaxEnt grammar, Hayes & Wilson (2008) created a learnability model known as the MaxEnt grammar tool. The method creates constraints and weights based on the sources provided by the incoming data, rather than being provided with constraints in advance. The calculation of Z is a concurrent issue since the set of potential candidates in OT is infinite, and it is difficult to calculate Z for an infinite set. The constraints are not given arbitrary weights in the learnability model of the MaxEnt grammar. Hayes & Wilson (2008) suggested that in order to learn the weights, the probabilities of the unobserved forms should be minimized while maximizing the probability of the observed data (P(D)). This ensures that all candidates carry a greater probability than all candidates who do not occur. It is also called Maximum Likelihood Estimation. Thus, P(D) is the product of the probabilities of each observed datum.

$$P(D) = \prod_{x \in \Omega} P * (x)$$

The machine-implemented learning method developed by Hayes & Wilson (2008) demonstrated how it is more straightforward than the majority of other phonological learning models now in use and can learn gradient phonotactics in a way that can be proven. In addition to better modeling phonotactic well-formedness than any other method, it handles important features of learning including hidden structure (structural ambiguity), free variation, etc. Using the current machine learning techniques, modeling the acquisition of grammar is made easier by its minimal requirement for data sampling.

Input	Output	Count	*T[-vce,-sg]syll	IDENT-IO	*[+spread glottis]
			cA	cB	cC
pap	paph	8	0	1	1
	pap	2	1	0	0
	phap	0	1	1	1
	phaph	0	0	3	1
calak	calak	2	1	0	0
	calakh	0	1	1	1

Table 15. Input data for aspiration

5.2. Implementing the Hayes and Wilson MaxEnt learner tool(2008)

The MaxEnt tool developed by Hayes and Wilson (2008) has mainly two functions as detailed below. The model is fed text data with the elaborated constraints. 1 stands for violation of certain constraints while 0 stands for satisfaction. Since the model cannot read the standard IPA transcription, they are written in English alphabetical form with vowels being specified differently.

The algorithm mainly –

1. Finds n possible phonotactic constraints for some corpus
2. Assigns weights for the constraints

Case 1: Aspiration

- Input data: Table 15
- Weights after optimization:

1. *T[-vce,-sg]syll - 10.74109
2. IDENT-IO - 6.23888
3. *[+spread glottis]- 3.323561

• Ranking based on probability: *T[-vce,-sg]syll » IDENT-IO » *[+spread glottis]

Case 2: Epenthesis

- Input data: Table 16
- Weights after optimization:

1. SONORITY- 4.436297962383264
2. ALIGN-L- 4.436297962383264
3. DEP-MF- 9.940723045863058

Input	Output	SONORITY	ALIGN-L	DEP-MF	ALIGN-R	DEP-IO	IDENT-IO
		cA	cB	cC	cD	cE	cF
skuTar	skuTar	4	1	0	0	0	0
	iskuTar	0	1	1	0	0	1
	sIkuTar	9	0	0	0	1	1
	sakuTar	1	0	0	1	0	0
spId	spId	5	1	0	0	0	0
	ispId	0	0	1	1	0	1
	sipId	8	0	0	0	1	1
	sApId	0	0	0	1	1	1

Table 16. Input data for epenthesis

4. IDENT-IO-0.31224672988385754

5. DEP-IO- 0.31224672988385754

6. ALIGN-L- 0.31224672988385754

Ranking based on probability: Dep (MF) » Align-L, Sonority » Ident-Io, Dep-IO, Align-R.

6. Conclusion

We can see that there are some issues while checking the learnability of phonological grammar. It is important to conduct an acoustic analysis of aspiration and epenthesis in Kashmiri loanwords, which is beyond the scope of the current study. Acoustic analysis is expected to provide more authentic and reliable data regarding the changes.

However, it is observed that Sonority occupies an impactful role in the acquisition of syllable structure in Kashmiri. The epenthesis, although handled differently in different languages, is crucial evidence of SSP as a vital principle of syllable formation. In terms of using models of learnability, we can see that RCD can converge into target grammar when presented with full structures, without any hidden representation. The ranking differs when the constraints are assigned numerical weights. This observation makes RCD less reliable. MaxEnt grammar appears to be more effective in considering gradient convergence.

Acknowledgements

I would like to thank my supervisor Prof. Shakuntala Mahanta for her valuable feedback on my work. I am also grateful to the anonymous reviewers of ConSOLE 31. Despite these reviews, I am sure this paper has its flaws and I myself am totally responsible for that. All errors are my own.

Sneha Ray Barman
Centre for Linguistic Science and Technology
Indian Institute of Technology Guwahati
raybarmansneha2@gmail.com

References

Albright, A. & Hayes (2011). Learning and learnability in phonology. Goldsmith, J. J. Riggle & A. C. L. Yu (eds.), *The handbook of phonological theory*, Oxford, Wiley-Blackwell, pp. 661–690.

Boersma, P. (1997). How we learn variation, optionality, and probability. *Institute of Phonetic Sciences of the University of Amsterdam (IFA) Proceedings 21*, pp. 43–58.

Boersma, P. & B. Hayes (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32:1, pp. 45–86.

Grierson, G. A. (1919). *Linguistic survey of India Vol. VIII: specimens of the Dardic or Pisacha languages (including Kashmiri)*. Superintendent Government Printing, Calcutta.

Hayes, B. (2007). The analysis of gradience in phonology: what are the right tools?. [Ms]. Stanford University. https://linguistics.ucla.edu/people/hayes/papers/HayesSlidesForStanfordGradienceWorkshop.pdf.

Hayes, B. & C. Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:3, pp. 379–440.

Itô, J. (1989). A prosodic theory of epenthesis. *Natural Language Linguistic Theory* 7:2, pp. 217–259.

Jarosz, G. (2019). Computational modeling of phonological learning. *Annual Review of Linguistics* 5, pp. 67–90.

Kager, R. (2004). *Optimality theory*. Cambridge Textbooks in Linguistics, Cambridge University Press.

Kang, H. (1996). English loanwords in Korean. *Studies in Phonetics, Phonology and Morphology* 2, pp. 21–47.

Kang, Y. (2003). Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology* 20:2, pp. 219–273.

Kar, S. (2013). Complex onsets in regional varieties of Bangla. *Language Sciences* 40, pp. 212–220.

Koul, O. (2005). *Studies in Kashmiri linguistics*. Indian Institute of Language Studies, Delhi.

Lee, H. (2000). English loanword phonology in Korean. *Language Research-Seoul* 37:1, pp. 177–202.

Lombardi, L. (2003). Markedness and the typology of epenthetic vowels. [Ms]. University of Maryland. https://roa.rutgers.edu/files/578-0203/578-0203-LOMBARDI-0-1.PDF.

Mahmood, R., Q. Hussain & A. Mahmood (2011). Phonological adaptations of English words borrowed into Punjabi. *European Journal of Social Sciences* 22:2, pp. 234–245.

Nagarajan, H. (2014). Constraints through the ages: loan words in Bangla. *The EFL Journal* 5:1, pp. 41–63.

Prince, A. & P. Smolensky (2004). *Optimality theory: constraint interaction in generative grammar*. Oxford, Blackwell.

Ramadoss, D. & K. G. Vijayakrishnan (2006). *Redefining the role of featural specifications in alignment*. [MA thesis]. The English and Foreign Languages University Hyderabad.

Shademan, S. (2002). *Epenthetic vowel harmony in Farsi*. [MA thesis]. Department of Linguistics, University of California, Los Angeles (UCLA).

Tesar, B. & P. Smolensky (1995). The learnability of optimality theory: An algorithm and some basic complexity results. [Ms]. John Hopkins University, Department of Computer Science. https://doi.org/10.7282/T34Q7SB7.

Tessier, A. M. (2017). Learnability and learning algorithms in phonology. *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.108.