October 2022

Restrictive Tier Induction

Seoyoung Kim
University of Massachusetts Amherst

RESTRICTIVE TIER INDUCTION

A Dissertation Presented

by

SEOYOUNG KIM

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2022

Department of Linguistics

RESTRICTIVE TIER INDUCTION

A Dissertation Presented

by

SEOYOUNG KIM

Approved as to style and content by:

Michael Becker, Chair

Gaja Jarosz, Member

Maria Gouskova, Member

Joe Pater, Department Chair
Department of Linguistics

DEDICATION

For mom, who taught me how to be resilient,

and dad, who taught me how to find humor during the toughest times.

ACKNOWLEDGMENTS

My deepest thanks go to my academic supervisor Michael Becker. He was the reason I initially went to Stony Brook, the reason I moved to UMass and stayed, and lastly the reason I am who I am as a doctor today. Thank you, Michael, for providing me with the invaluable insight that brought clarity to my chaotic thoughts, for always guiding me to the next steps, for being accessible and fun, and also for being supportive and patient with me. I also must thank Gaja Jarosz. Working with her opened up a whole new set of opportunities in the area of computational linguistics. She would always listen closely to my learning simulation problems that agonized me for weeks and give me on the spot solutions. I first thought of her as my academic mentor, but after working with her for 3 years, she has become my absolute inspiration in so many areas of life! Gaja, having a powerful and understanding female figure like you has been incredibly inspiring during my time at UMass. Maria Gouskova's work was the initial inspiration for this dissertation. She kindly accepted the invitation to be on my dissertation committee and then gave me insightful comments and helpful feedback. Without her work, resources, and generosity, my dissertation would not have taken the shape that it has today.

Transitioning to a different program in my third year was a unique experience. Naturally, I was sad to leave Stony Brook but now I have come to view my experience somewhat differently. I feel very grateful to have known great people from more than one place. I still talk to many of you at Stony Brook. My biggest thanks of course go to my cohort: Khanin and Andrija. Friends, we all made it! I also thank Hyunah, Sophie, Alëna, Aniello, and Ayla for making my time in Long Island surprisingly missable. Fitting into a new school as a third year *during a pandemic* was not easy.

However, I was extremely fortunate to have Duygu and Anissa just a few steps away, who accepted me into their sisterhood with sincere hearts as if I always had belonged in your cohort. My special thanks also go to Andrew for always being there for me, academically and non-academically, not only figuratively but also literally. Occasional parties with Shay, Max and Katie, Maggie, Özge, and Jed were such fun! Guys, I can't believe that it's time for me to leave when I feel most close to all of you. Lastly, I thank David and Potato for constantly reminding that there is life beyond linguistics and for letting me know that cuteness can cure nearly all problems.

ABSTRACT

RESTRICTIVE TIER INDUCTION

SEPTEMBER 2022

SEOYOUNG KIM

B.A., YONSEI UNIVERSITY

M.A., SEOUL NATIONAL UNIVERSITY

M.A., STONY BROOK UNIVERSITY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Michael Becker

This dissertation proposes the Restrictive Tier Learner, which automatically induces only the tiers that are absolutely necessary in capturing phonological long-distance dependencies. The core of my learner is the addition of an extra evaluation step to the existing Inductive Projection Learner (Gouskova and Gallagher 2020), where the necessity and accuracy of the candidate tiers are determined.

An important building block of my learner is a typological observation, namely the dichotomy between trigram-bound and unbounded patterns. The fact that this dichotomy is attested in both consonant interactions and vowel interactions allows for a unified approach to be used. Another important piece of information is that only unboundedness implies trigram-boundedness, and not vice versa. These typological observations together shed light on the critical role of trigrams in phonological learning. The premise that there is no other distance at which a restriction holds than

these two lets us safely assume that searching only up to trigrams might actually be a near-exhaustive search for local interactions. On top of that, the fact that interaction beyond a trigram window, which we need tiers for, always implies interaction within a trigram window guarantees that all necessary tiers can be discovered by looking at trigram constraints. Hence, a learner can confidently search up to trigrams for local interactions and expand its search for non-local ones from the discovered trigrams.

I present several case studies to test the abilities of the Restrictive Tier Learner in capturing various long-distance dependencies that are attested in natural languages. The current version of the learner maintains all the strengths of the previous learning algorithms while showing improved performance in critical cases.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

> The result shows the sufficiency of vowel projections for the task at hand, but to demonstrate necessity would be a long-term project.
>
> −Hayes and Wilson (2008)

This dissertation proposes a learner that automatically induces only the tiers that are absolutely necessary in capturing phonological long-distance dependencies. I exploit a typological observation as a heuristic to aid in determining the necessity and accuracy of the candidate tiers. In this chapter, I introduce the motivations of the dissertation.

1.1 Long-distance dependencies and phonological tiers

Nonlocal phonological interactions such as vowel harmony and long-distance consonant assimilation/dissimilation often hold at arbitrary distances. Shown below is Navajo, a language with sibilant harmony in which the anteriority of sibilants should match within roots; in (1a), the root only contains [−anterior] sibilants. Harmony does not generally hold within a word but some prefixes alternate, assimilating to the anteriority of the root sibilant; in (1b)−(1c), the perfective prefix /si-/ surfaces as [ʃi-] if the root contain a [−anterior] sibilant and faithfully otherwise. Moreover, harmony is stronger transvocalically and is optional with more intervening material (Sapir and Hoijer 1967, Berkson 2010, Gallagher 2020). In (1d), the realization of /t͡ʃ/ can either be faithful or harmonizing to the word-final [s], indicating that the application of harmony is optional at further distances.

(1) Navajo sibilant harmony (Sapir and Hoijer 1967)

a. wóʧ'oʃí — 'rotten tooth'

b. si-tį́ — 'he is lying'

c. ʃi-tééʒ — 'they two are lying'

d. ʔa-ʧii-ɬ-tááʃ ∼ ʔa-tsii-ɬ-táás — 'he bends things'

Capturing this pattern requires not only a markedness constraint such as *[α anterior][ ][β anterior], meaning that posterior and anterior sibilants cannot co-occur over a segment, but also requires *[α anterior][ ][ ][β anterior], *[α anterior][ ][ ][ ][β anterior], and so forth until it reaches the length of the longest word in this language. Since the word length is unbounded and the number of constraints increases accordingly with the possible word length, this kind of description requires an infinite hypothesis space and makes exhaustive search impossible (Hayes and Wilson 2008, Gouskova and Gallagher 2020). For this reason, nonlocal phonological interactions motivate representations that include only the interacting segments, namely a *tier* on which only the interacting segments are visible. For example, the Navajo case above can be simply described as *[α anterior][β anterior] on a separate tier on which only sibilants are visible.

Tier-based representations enable the description of long-distance dependencies and there is a near consensus among phonologists that we need tiers to describe such dependencies; but can a computational model find tiers automatically?

Hayes and Wilson (2008) offers a learner that induces markedness constraints against unattested or underattested structures. While their learner successfully captures local interactions, generalizing long-distance dependencies requires a projection of a tier that is pre-defined by the analyst. For example, they tested their learner on the Shona vowel harmony system where only vowels of the string participate. They reported that capturing the Shona vowel harmony pattern is impossible unless a separate vocalic representation is given in advance to their learner. And crucially, the model cannot learn tiers on its own.

Gouskova and Gallagher (2020) introduced a learning algorithm, which they call the *Inductive Projection Learner*, where long distance interactions can be learned without having to predefine tiers. They argue that tiers can be induced from the phonotactic properties of a language that are observable as local trigrams. For instance, Gallagher (2020) reports a learning simulation on Navajo roots; their learner induced a [+strident] tier automatically based on evidence observable in local trigrams (e.g., CVC). Contrary to the learner of Hayes and Wilson (2008) where tiers need to be defined by the analyst and supplied in advance, Gouskova and Gallagher (2020) restrict the power of the analyst and automatically induce tiers that are evidenced in the input data.

1.2 Goal of the dissertation

Hayes and Wilson (2008) commented on their learning simulation of Shona vowel harmony that their "result shows the sufficiency of vowel projections for the task at hand, but to demonstrate necessity would be a long-term project". As a first step of the said long-term project, I explore the necessity of phonological tiers from two different standpoints in the dissertation.

First, I investigate whether having a phonological tier is actually necessary in various language patterns that have been attested. The learner of Gouskova and Gallagher (2020) interprets local trigrams such as *X[ ]Y as evidence that X and Y interact nonlocally and that a tier projection based on X and Y is therefore necessary. As I will show in the later chapters, this is not always true. Some restrictions hold only within a bounded window, indicating that the existence of a local trigram does not always guarantee that the pattern holds at arbitrary distances. A tier projection based on a local trigram leads to projecting unnecessary, and often too many, tiers.

Secondly, I investigate the necessity of segments that are projected on the tier. As mentioned above, the algorithm of Gouskova and Gallagher (2020) projects a tier based on the existence of local trigrams. However, trigrams might not be sufficient to capture the full range of patterns because some patterns no longer hold when a specific segment (i.e., blocker) intervenes. Hence, such strings with opaque interveners can be mischaracterized on the tier, if the tier is projected only based on a local trigram.

A tier inducing algorithm, therefore, should have the ability to distinguish bounded patterns from unbounded ones, correctly discerning whether a tier projection is actually required. Furthermore, a tier inducing algorithm also should ideally have the ability to include necessary segments while excluding unnecessary ones. These are the goals that I try to achieve by introducing a new tier-inducing algorithm in the dissertation. The overview structure of my learner is shown below.

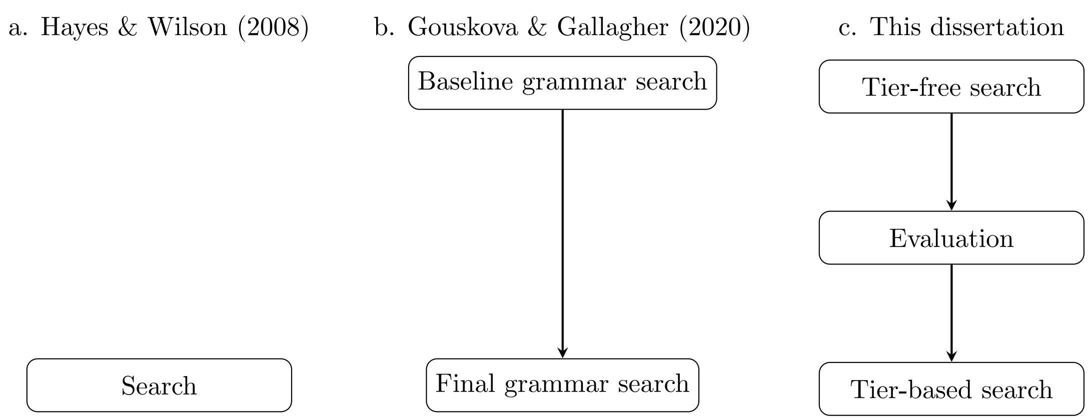

Figure 1. The overview structure the learner

Shown in Figure 1(a) is the Hayes and Wilson learner where tiers need to be manually provided by the analyst. To this algorithm, Gouskova and Gallagher (2020) add a procedure in which tiers can be automatically discovered, shown in Figure 1(b). I add an extra step, called evaluation, between the baseline grammar search and final grammar search, as illustrated in Figure 1(c). Unlike the learner of Gouskova and Gallagher where the projection of a tier is automatic from each trigram, my learner goes through *evaluation* before proceeding to a tier projection, where the necessity of candidate tiers is evaluated.

1.3 Structure of the dissertation

The rest of the dissertation is organized as follows. Chapter 2 gives an overview of the typology of long-distance dependencies in phonology. The typology of locality is an important basis of the dissertation because I derive a heuristic from the typological observation and use it to build a learning algorithm.

Chapter 3 introduces the Restrictive Tier Learner, the learning algorithm that I propose in the dissertation. I first introduce the Inductive Projection Learner (Gouskova and Gallagher 2020), which is the base learning algorithm of my learner. I also lay out the predictions the Inductive Projection Learner makes; importantly, I show that their learner projects an unnecessary tier when the pattern is merely trigram-bound and cannot discover blockers, leading the grammar to mischaracterize words with opaque segments (blockers). Subsequently, I illustrate the structure of the Restrictive Tier Learner in-depth, along with providing relevant real-life language examples.

Chapter 4 presents the learning simulation results produced by my proposed algorithm; I will also provide learning results of the Inductive Projection Learner when comparison is necessary. Overall, my learner does not deteriorate what the previous learning algorithms do while showing improved performance on some patterns.

Chapter 5 discusses remaining issues around my learner, specifically the assumptions and predictions that my learner makes. I also characterize limitations and suggest possible directions that future study can pursue. Chapter 6 summarizes and concludes the dissertation.

CHAPTER 2

BACKGROUNDS: THE TYPOLOGY OF LOCALITY

This chapter gives an overview of the typology of long-distance dependencies in phonology. There are various dimensions that can be considered in introducing the typology, such as locality, the presence of decay and also the cause of decay, the type of interaction (assimilatory or dissimilatory), and whether it is a dynamic alternation process or a static morpheme structure rule.

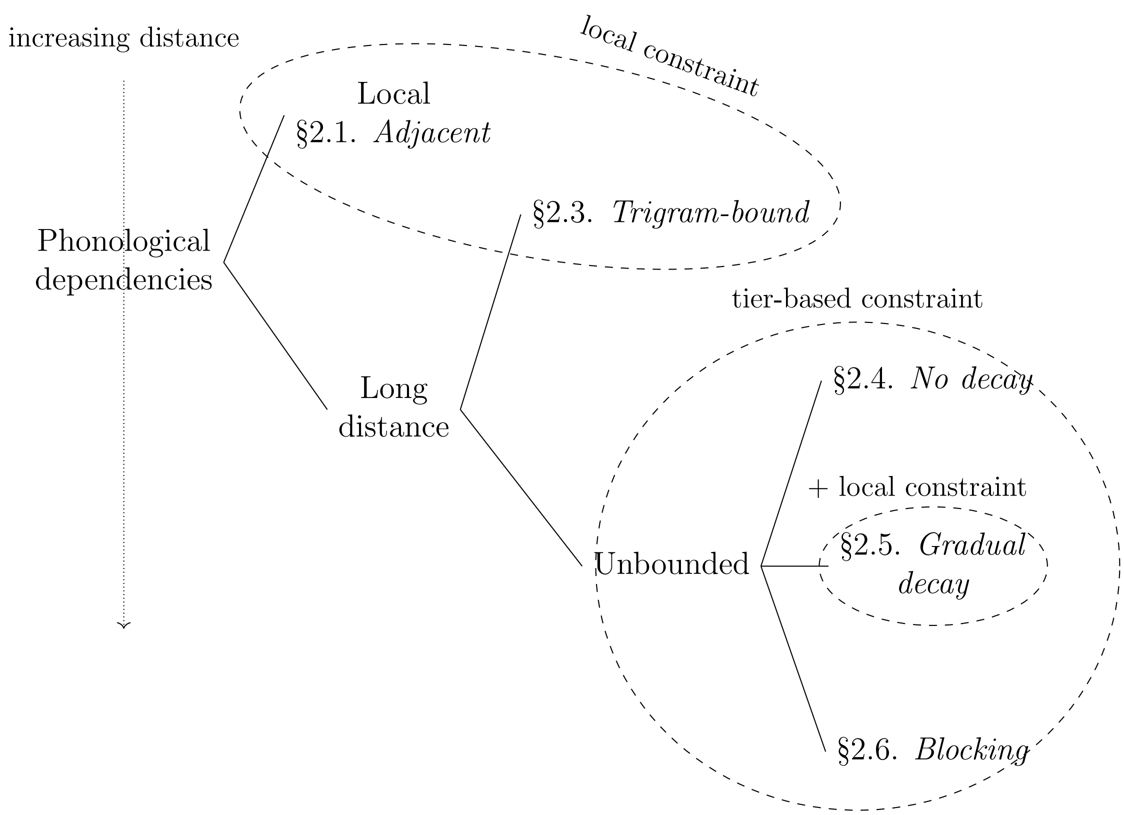

Figure 2. The typology of locality in phonological dependencies

I organize the chapter by locality, specifically in order of decreasing proximity, as shown in Figure 2: adjacent, trigram-bound, unbounded. The other properties that are relevant to the other dimensions will be briefly mentioned for each locality type.

In §2.1, I present cases where adjacent segments interact, such as in voicing or place assimilation in consonant clusters. In the rest of the chapter, I introduce long-distance dependencies. Long-distance here means not immediately adjacent; or put differently, that there is at least one intervener between the interacting segments. The focus of this dissertation is the analysis of these patterns from the perspective of phonotactic learning. I start the description of the typology by introducing the core observation in §2.2: the locality dichotomy between trigram-boundedness and unboundedness (Martin 2005, Walker et al. 2006, McMullin 2016, and Gallagher 2020). I then briefly discuss how this dichotomous nature can benefit learning and improve computational efficiency. I then discuss each pattern more in depth. In §2.3, I introduce trigram-bound patterns; these patterns are non-local in the sense that there is one segment that intervenes, but the locality dichotomy observed in the typology as well as the universal preference toward simple syllable structures (e.g., CV.C and V.CV) allow computationally analyzing these patterns as local trigrams. Throughout §2.4−§2.6, I introduce three sub-categories of unbounded patterns, which require the grammar to have access to tier-based representations. The three categories differ in whether the restriction weakens, and, if it does, whether the decay is caused by the distance between the interacting segments or/and the identity of the segments that intervene. I summarize the chapter in §2.7.

2.1 Interactions between adjacent segments

Numerous languages require that neighboring sounds be somewhat similar to each other. Famous examples come from English, in which the negative prefix /ɪn-/ assimilates to the following segment, and more specifically to the place of articulation of the initial consonant of the root, as shown in (2).

(2) English place assimilation: prefixal /n/ assimilates to the following consonant

a. i[n-d]ecisive

b. i[m-p]ossible

c. i[ŋ-k]ongruent

It is also very common that many languages do not allow a sequence of vowels that are not identical (vowel hiatus). While non-identical vowel sequences can be repaired by deleting one of the vowels or inserting a consonant in between, some languages utilize assimilation as a repair strategy. As shown in (3), Yoruba exhibits regressive vowel assimilation in which the preceding vowel assimilates to the following one. As shown in (3), the final vowel [e] in [owe] assimilates to the immediately following [a] in the second example.

(3) Yoruba vowel assimilation (Welmers 1973)

a. [owe] 'money'

b. [owa-ade] 'Ade's money'

The assimilation of adjacent segments is conceived as natural and intuitive, as it is very clearly supported by phonetic motivation; it saves gestural movements and thus improves ease of articulation. Although it is less common, local dissimilation also has been attested. Given in (4) is Ainu, a language spoken in Hokkaido of Japan, has a dissimilatory phenomenon which turns an /-ɾɾ-/ sequence into [-nɾ-]. As seen in the examples, the word-final /ɾ/ stays faithful unless there is another /ɾ/ that is immediately following it; in (4b), there is no dissimilation because the second /ɾ/ is too far away from the first /ɾ/. Here, the dissimilation process gives prominence to the word boundary.

(4) Ainu rhotic dissimilation (Shibatani 1990, Suzuki 1998): /-ɾɾ-/ → [-nɾ-]

a. /kukoɾ ɾusuy/ [kukon ɾusuy] 'I want to have (something)'

b. /kukoɾ kuɾ/ [kukoɾ kuɾ] 'my husband'

Adjacent assimilation or dissimilation is very commonly observed across many languages and it can be can be easily captured in various frameworks. In formal language theoretic approaches (Heinz 2010), adjacent patterns belong to the class of Strictly Local (SL) languages wherein local dependencies are enforced by local constraints, such as a local bigram *[ɾɾ]. Similarly, constraint-based frameworks can capture these patterns using local constraints. In a SPE-style analysis, adjacent processes can be expressed as a phonological rule (Chomsky and Halle 1968), as in /ɾ/ → [n]/__ ɾ.

2.2 The dichotomy: trigram-bound vs. unbounded

This section introduces phonological dependencies that are long-distance, in a sense that segments interact across at least one another segment. McMullin (2016) argues that, at least for consonant harmony, there is a robust dichotomy between dependencies that apply only within a bounded ...CVC... window and unbounded patterns, which hold in all ...C...C... contexts, and there is no other type of restriction on distance that is attested. He demonstrates this dichotomy by showing that natural languages with sibilant harmony or nasal harmony all fall into either pattern. The languages in (5)−(6) both exhibit restrictions in which the anteriority of sibilants must match within a certain domain.

(5) Aari (Hayward 1990): /-s/ → [-ʃ] with any preceding [−ant] sibilant

a.	/baʔ-s-e/	baʔ-**s**-e	‘he brought’
b.	/ʧ’a̤:q-s-it/	**ʧ**’a̤:q**ʃ**it	‘I swore’
c.	/ʒa̤:g-er-s-e/	**ʒ**a̤:ger**ʃ**e	‘it was sewn’

(6) Koyra (Hayward 1982): /-os:o/ → [oʃ:o] with a sibilant within a trigram window

a.	/tim-d-os:o/	tindo**s**:o	‘he got wet’
b.	/paʧ-d-os:o/	pa**ʧ**:o**ʃ**:o	‘it became less’
c.	/ʃod-d-o**s**:o/	**ʃ**od:-o**s**:o	‘he uprooted’

The examples given in (5) show that the suffixal /s/ must surface as [ʃ] if there is a preceding [−anterior] sibilant anywhere in the word; in the two last words, the suffix harmonizes to the anteriority of the preceding sibilant regardless of their distance. In the first example, the suffix occurs faithfully as [s] because the root does not have any [−anterior] sibilant that the suffixal /s/ should harmonize to. Koyra, shown in (6), also exhibits similar restrictions but harmony is required only if the trigger of the harmony, a [−anterior] sibilant, is present within a trigram window. In the second word, the suffixal [s] surfaces as a [ʃ] due to the root /ʧ/ in the immediately preceding syllable; the two sibilants that occur across a single vowel should agree in their anteriority. In the last example, however, the suffixal [s] still stays faithful regardless of the root /ʧ/ because there are more segments that intervene between the two sibilants.

A similar typological dichotomy is observed in vowel interactions. McCollum (2019) points out that there are only two types of attested vowel harmony pattern in terms of their iterativity; vowel harmony can be either non-iterative where only a single vowel is harmonized within a given domain ...VCV... or fully iterative throughout the domain ...V...V...V... Crucially, he further adds that there is very little evidence for a pattern where *n* vowels may harmonize within a given domain. For example, shown in (7) is Karajá (Ribeiro 2002). In this language, a morpheme containing a [+ATR] vowel can trigger regressive harmony, regardless of its morphological affiliation. In the example, the [+ATR] vowels of the imperative particle (/ikuđi/) turns all preceding [−ATR] vowels into [+ATR]; the harmony process here is iterative (V...V...V).

(7) Karajá (Ribeiro 2002): regressive iterative [+ATR] harmony

/b-ɛ-ɗɛhɛ-ikuđi-hɛ/	[beɗeheikunihɛ]	'2ND-INTR-look-IMPER-EMPH'

The language shown in (8) exhibits non-iterative round harmony in which only one vowel harmonizes to the triggering [+round] vowel.

(8) Kazakh (McCollum and Kavitskaya 2018): non-iterative round harmony

a.	/mojən-də/	m**o**j**ʊ**ndə	'neck-ACC'
b.	/kino-m-əz-dəŋ/	kin**o**-m-**ʊ**z-dəŋ	'movie-POSS.1-PL-GEN'

The examples laid out so far show assimilatory effects between consonants and between vowels, and also how there is a dichotomy in terms of locality involved: consonant harmony and vowel harmony can be either bounded to a trigram-window or unbounded. The same dichotomy of locality is found in consonant dissimilations and vowel dissimilations as well. Suzuki (1998) investigates the typology of dissimilation in various languages. In terms of locality, he distinguishes syllable adjacency as in CV.C, single consonant adjacency as in V.CV, and unboundedness; syllable adjacency and single consonant adjacency essentially refer to trigram-boundedness.

Yimas, given in (9), is a Papuan language spoken in New Guinea, exhibits a transvocalic dissimilation of alveolar taps. As seen in (9b), the inchoative suffix /-aɾa/ surfaces as [-ata] with a root-final alveolar tap /ɾ/ but the suffix surfaces faithfully with a farther /ɾ/ within a root, as seen in (9c). The example of consonant dissimilation which holds at unbounded distances is given in (16). Latin is known to have a dissimilatory phenomenon in which the adjectival suffix /-alis/ turns into [-aris] with a preceding /l/ in the root. As seen in (16a), the suffix surfaces faithfully with an absence of an /l/ in the root. The suffix surfaces as a dissimilated form [-aris] after a root that contains an /l/, shown in the three other examples. In fact, the Latin pattern becomes more complicated with extra factors playing a role; the likelihood of the dissimilation depends on the distance between the interacting liquids as well as the identity of interveners. I will introduce the other aspects of the Latin dissimilation later in §2.5 more in depth, but the point that is made about this pattern for now is the fact that it can hold at arbitrary distances (unbounded).

(9) Yimas: /-aɾa/ → [-ata] with root-final [ɾ] (Shibatani 1990, Suzuki 1998)

a.	/pak-aɾa/	pak-aɾa	'break open'
b.	/apɾ-aɾa/	apɾ-ata	'open spread'
c.	/kkɾak-aɾa/	kkɾak-aɾa	'loosen'

(10) Latin: /-alis/ → [-aris] with any preceding /l/ (Zymet 2014, Bennett 2013, Stanton 2016a)

a.	[kib-alis]	'of food'
b.	[so**l**-aris]	'solar'
c.	[lana-ris]	'of wool'
d.	[lapida-ris]	'of rocks'

Non-local vowel dissimilation is also attested in both transconsonantal (V.CV) configuration and unbounded window. Examples in (11) are from Kera, in which the low vowel /a/ turns into a [ə] with a following [a] within a trigram window. Given in (12) is Malagasy, in which the passive imperative suffix /-u/ turns into an [i] in the presence of another /u/ anywhere in the root. In (12a), the suffixal vowel surfaces faithfully whereas it surfaces as a dissimilated form /-i/ in the three other examples. Vowel dissimilation in Malagasy peters out with the increasing number of interveners, just as in the Latin case, but the point here is, again, that it can hold outside the trigram window and unboundedly.

(11) Kera (Suzuki 1998): low vowel /a/ → [ə] with a following [a] within a trigram window

a. ba ‘not’

b. pa ‘again’

c. bə-pa ‘no more’

d. balna-n ‘wanted me’ (*[bəlna-n])

(12) Malagasy (Zymet 2014): suffixal /-u/ → [-i] in the presence of /u/ in the root

a. /bata-u/ [batau] ‘lift’

b. /tuv-u/ [tuvi] ‘fulfill’

c. /tuda-u/ [tudai] ‘prevent’

d. /gurabah-u/ [gurabahi] ‘spluttering’

The information provided in the chapter so far is briefly summarized in Table 1x; for each locality type, adjacent, trigram-bound, and unbounded, assimilation and dissimilation patterns are both attested in consonant and vowel interactions. Importantly, the locality dichotomy of non-local patterns, the distinction between trigram-boundedness and unboundedness is robust in both consonant and vowel assimilation and dissimilation. As I will show later in Chapter 3, this resemblance allows utilizing a unified approach to automatically inducing tiers for both consonant and vowel interactions.

	Consonant	Vowel
adjacent	CC	VV
bounded	CVC	VCV
unbounded	C...C	V...V

Table 1. The typology of locality in consonant and vowel interactions

Related to this dichotomy, Walker (2000), Hansson (2010a), Finley (2011), and McMullin and Hansson (2019) point out that there is an asymmetric implicational relation between these two patterns; as depicted in Figure 3, interaction at beyond transvocalic distances entails interaction in transvocalic contexts, but not vice versa. Therefore, it can be summarized that a consonantal harmony observable as a local trigram *guaranteed* to generalize unboundedly if it holds over another consonant. Similarly, a vowel harmony captured as a local trigram is *guaranteed* to generalize beyond the trigram window and unboundedly if it holds over another vowel.

$$Unbounded \underset{\not\Longleftarrow}{\Longrightarrow} Trigram - bound$$

Figure 3. Asymmetric implicational relation

The significance of this dichotomy is most strongly recognized from a learnability perspective. The premise that there is no other non-local distance at which a restriction holds than these two lets us safely assume that searching only up to only trigrams, which is quite manageable, might actually be a near-exhaustive search for adjacent and trigram-bound patterns. Importantly, although being non-local by their nature, transvocalic and transconsonantal patterns can be found via local trigrams. On top of that, the fact that interaction beyond a trigram window, which we need tiers for, always implies interaction within a trigram window guarantees that all necessary tiers can be discovered by only looking at trigram constraints. This strict dichotomy and the asymmetry can substantially benefit learning since it limits the search space without sacrificing accuracy; a learner can confidently search up to trigrams for local interactions and expand its search for non-local ones from the discovered trigrams.

The asymmetry given in Figure 3 is also in line with the prediction that is made by the Proximity Hierarchy (Suzuki 1998), in which some markedness constraint is incorporated with the universal hierarchy on the size of the intervening material. As can be seen in (13), structures with a certain co-occurrence X...X (where X is a consonant) can be penalized by a more dominant constraint if they are closer to each other. Since *X-V-X (transvocalic) dominates *X-∞-X (unbounded) in the hierarchy, if a language bans a certain structure at an arbitrary distance, it should also rule out the same structure over a single vowel. As mentioned above, adjacent patterns are not the direct focus of the dissertation but it is worth noting that the implicational asymmetry generalizes to the adjacent patterns as well; according to the hierarchy, if a language bans a certain structure within a transvocalic window or within unbounded distance, the same structure should also be banned when it is adjacent, since *XX (adjacent) dominates *X-V-X (transvocalic) and *X-∞-X (unbounded). For example, Japanese does not tolerate more than one voiced obstruent within a word (Itô and Mester 1986). The proximity hierarchy predicts that the language should also ban transvocalic or adjacent co-occurrence of two voiced obstruent. As predicted, Japanese indeed does not allow voiced obstruent geminates (e.g., *[bb]) as well as co-occurrence over a vowel (e.g., *[badu]).

(13) Markedness constraint + Proximity Hierarchy

$$*\text{X...X} = \{*\text{XX} \gg *\text{X-V-X} \gg ... \gg *\text{X-}\sigma\text{-X} \gg ... \gg *\text{X-}\infty\text{-X}\}$$

For the rest of this chapter, building on this typological dichotomy, I lay out four different types of long-distance interactions in phonology: trigram-bound in §2.3, unbounded no decay in §2.4, unbounded gradual decay in §2.5, and unbounded selective decay in §2.6.

2.3 Transvocalic and transconsonantal patterns

Transvocalic or trasconsonantal interaction refers to cases where the restriction is strictly bounded to a trigram window and immediately shuts off as soon as extra segments intervene. This pattern has been attested in numerous Bantu languages as a form of consonantal interactions (Hansson 2010a, McMullin 2016). For example, in Koyra, the perfective suffix /-os:o/ surfaces as [-oʃ:o] if a [−anterior] consonant precedes within a transvocalic window, and otherwise surfaces faithfully (Hayward 1982, Hansson 2010a, McMullin 2016). There is also Lamba, a Bantu language spoken in Zambia and Congo, given in (14). In Lamba, the perfective suffix /-ile/ surfaces as [-ine] if a nasal consonant precedes across a single vowel, and otherwise surfaces faithfully (Odden 1994, Hansson 2010a, and McMullin 2016).

(14) Lamba transvocalic nasal harmony (Odden 1994)

a.	/-pat-ile/	[-patile]	'scolded (perf.)'
b.	/-u:m-ile/	[-u:mine], *[-u:mile]	'dried (perf.)'
c.	/-mas-ile/	[-masile], *[-masine]	'plastered (perf.)'

In example (14a), the suffix surfaces faithfully as [-ile] because there is no nasal in the root which can trigger the suffix alternation. In (14b), the suffix surfaces as a nasalized form [-ine] because of the stem-final nasal *m*. Notably, there is only a single vowel between the nasal trigger and the target in the suffix, letting the transvocalic nasal harmony happen. In (14c), the alternation is not triggered because the stem *m* is outside the trigram window, demonstrating that the nasal harmony holds only over a single vowel.

Similar patterns are found in vowel interactions, in which dependencies between vowels hold only across a single consonant and do not hold outside of a trigram window (e.g., VCV). The most commonly attested pattern that fits into this category would be non-iterative harmony, found in numerous Turkic languages (McCollum and Kavitskaya 2018, McCollum 2019, McCollum and Kavitskaya 2022). For instance, in Kazakh as previously presented in (8), underlying [+round] vowels trigger harmony exclusively on the following syllable and not any further; in forms like [mojʊn-də] (*[mojən-də], *[mojʊn-dʊ]) 'neck-ACC' and [kino-m-ʊz-dəŋ] (*[kino-m-əz-dəŋ], *[kino-m-ʊz-dʊŋ]) 'movie-POSS.1-PL-GEN', only a single vowel immediately after a [+round] vowel is harmonized.

The patterns introduced in the current section are ones that are strictly bounded to a trigram window, such as VCV or CVC. This allows that transvocalic and transconsonatal patterns, although not categorized as "adjacent", analyzed via local mechanisms, such as local trigrams. In languages that allow consonant clusters or vowel hiatus, dependencies can also hold over multiple consonants or multiple vowels. From the perspective of typological description, the distinction between unbounded versus transvocalic/transconsonantal still holds; the fact that the harmony holds over a consonant cluster or a vowel cluster does not void the name *transconsonantal* (VCCV) and *transvocalic* (CVVC). From the perspective of computational learning, languages with complex syllable structures require some extra complexity; for example, if iterative vowel harmony holds over a consonant cluster, as in (VC.CV), it can no longer be captured as a local trigram. As Gouskova and Gallagher (2020) point out, however, strings with CV structure are universally more frequent even in these languages that allow consonant clusters or vowel hiatus, which still allows the discovery of trigram constraints. Thus, trigrams can still be a good starting point for inducing tiers regardless of what syllable structures a language allows. I will continue this line of discussion in §5.4.

Lastly, while transvocalic and transconsonatal patterns were conventionally analyzed with a focus on their boundedness, these bounded patterns can also be seen as an abrupt distance-based decay; the transvocalic restriction such as *[*nm*]...*l* immediately decays all the way if extra material intervenes and therefore a sequence of a nasal and a liquid is no longer bad at a farther distance, as in [-**m**asi**l**e]. Similarly, the constraint that encourages round harmony, such as *[+round]...[−round] can be seen as immediately losing its importance outside a certain bounded window in cases of non-iterative harmony.

2.4 Unbounded No Decay

In the previous section §2.3, I discussed the attested language patterns that belong to the *trigram-bound* category of Figure 3. In the remaining sections of this chapter, I introduce three other patterns that can all be fitted into the category of *unbounded*: Unbounded no decay, distance decay, and selective decay. While these three patterns are unbounded in the sense that there is no available evidence of being bounded, they differ in the presence and the motive of decay.

The first case of unbounded pattern would be thought of as a true unboundedness. *Unbounded No Decay* refers to cases where a restriction is at work unboundedly, at the same strength without decaying, regardless of the identity of interveners or the number of interveners. Categorical laryngeal phonotactics in Quechua, in which ejectives and aspirates may not follow plain stops within a word, can be an example of this case; it is a categorical and inviolable restriction, meaning that other factors, such as distance between the interacting stops or the existence of a morpheme boundary, do not weaken the restriction. For example, *[kap'i] and *[kasp'i] are equally ungrammatical although the plain stop *k* and the ejective *p'* are farther apart in *[kasp'i] (Gouskova and Gallagher 2020). Although these restrictions in Quechua are known to be categorical and inviolable, meaning that other factors, such as the number of interveners, do not weaken the restrictions, the actual unboundedness of this restriction cannot be checked, because the canonical root of the language is only disyllabic (Gouskova and Gallagher 2020) and the language lacks sufficiently long words. In fact, as Zymet (2014) has already suspected, all long-distance processes could be subject to distance decay but it is impossible to find data in some languages where segments are sufficiently far apart, partially due to the general dearth of long words (Stanton 2016b). Gallagher (2016) reports that there are only twelve trisyllabic roots with an ejective as the onset of the third syllable in the entire dictionary of Quechua, such as in *humint'a*. While none of these twelve words violate the above-mentioned restriction, she concludes that the evidence for the distance effect in Quechua is small. In a nutshell, there is no clear-cut evidence of the unboundedness and the distance effect in Quechua. Put differently, there can be a possibility that the restriction decays with distance or does not generalize unboundedly. However, I abstract away from these possibilities and still consider cases like Quechua to be the closest natural-language example of the scenario *Unbounded No Decay*, based on the fact that the restriction is unbounded and does not decay within the possible word length.

The corresponding vowel interaction pattern to the *unbounded no decay* consonantal interactions could be fully iterative harmony. Shown in (15) is Kinande.

(15) Kinande ATR harmony (Archangeli and Pulleyblank 1994, Cole and Kisseberth 1994)

a.	/tU-ka-kI-lim-a/	tukakilima	'we exterminate it'
b.	/tU-ka-kI-huk-a/	tukakihuka	'we cook it'
c.	/tU-ka-kI-lɪm-a/	tʊkakɪlɪma	'we cultivate it'
d.	/E-rI-hʊm-a/	ɛrɪhʊma	'to beat'

In this language, the ATR specification of the root (underlined) vowel spreads regressively to the prefixes all the way up to beginning of the word, (Archangeli and Pulleyblank 1994, Cole and Kisseberth 1994). The low vowel /a/ is transparent to vowel harmony, meaning that it neither participates in harmony nor blocks spreading of the ATR specification of the root vowel.

2.5 Unbounded gradual decay

The other two unbounded patterns both exhibit a decay of the restriction but the decay could be based on either the distance between the interacting segments or the identity of the intervening segment. The first unbounded decay pattern is based on the distance between the interacting segments; cases where a phonological restriction gradually peters out when the interacting segments are separated by more material. Take the examples from Latin in (16).

(16) Latin unbounded gradual decay in *l* dissimilation

a. [so**l**-aris] (*[so**l**-a**l**is]) 'solar'

b. [wu**l**g-a$_\sigma$lis] 'of wheat'

c. [di**l**uwi$_\sigma$-a$_\sigma$lis] 'of floods'

d. [so**l**sti$_\sigma$ti$_\sigma$-a$_\sigma$lis] 'of the summer solstice'

e. [**l**argi$_\sigma$ti$_\sigma$o$_\sigma$n-a$_\sigma$lis] 'belonging to imperial treasury'

In Latin, the realization of the adjectival suffix /-alis/ and the nominal suffix /-al/ largely depends on the presence of *l* in the root (Cser 2007, Cser 2010). With roots that contain no *l*, the /-alis/ and /-al/ suffix show up faithfully as [-alis] and [-al], as in [nav-alis] 'naval'. The /-alis/ and /-al/ suffixes surface as their dissimilated forms [-aris] and [-ar] after a root that contains /l/, as in (37a). This pattern has been traditionally analyzed as dissimilation for the feature [+lateral], where an underlying sequence of *l...l* is mapped to a surface form *l...r*. As will be explained with more detail in later chapters, the *l* dissimilation is also known to be a gradient process whose likelihood depends on the distance between the stem *l* and the suffixal *l*. Zymet (2014) and Stanton (2016a) both report that the dissimilation probability significantly decreases as the two [l]s are farther away from each other, while two *l*s barely co-occur when they are in adjacent syllables. Thus, for example, forms like [sol-aris] in (37a) will almost never surface as *[sol-alis], where two *l*s are onsets of adjacent syllables.

Similar patterns are found in vowel interactions; in some languages, the likelihood of vowel harmony decreases as the number of intervening syllables increases. In Hungarian, the dative suffix shows a two-way alternation in backness, which appears as [-nɔk] or [-nɛk], depending on the frontness of the root vowel (Hayes and Londe 2006, Hayes et al. 2009). Examples are given in (17). If the suffix attaches to a root that ends in a [+back] vowel, as in (17a), the suffix surfaces as its back form. If the root has a back vowel elsewhere in the word, both suffixes can be attached; but the likelihood of choosing the back suffix depends on the distance between the root back vowel and the suffix (Hayes and Londe 2006, Zymet 2014). Compared to the cases like (17a) in which the back suffix is almost always chosen, back harmony applies less reliably when they are separated by one neutral vowel (17b), and even less reliably when they are separated by two neutral vowels, as in (17c).

(17) Hungarian (Hayes et al. 2009)

a. [biːroː-nɔk] 'judge-DAT'

b. [ɔrzeːn-nɔk], [ɔrzeːn-nɛk] 'arsenic-DAT'

c. [poezis-nɔk], [poezis-nɛk] 'poetry-DAT'

Distance-based decay of vowel dissimilation is also attested. In Malagasy, which was previously presented in (12), dissimilation is less likely as the number of intervening segment increases (Zymet 2014).

From the perspective of distance-based decay, the patterns that belong to this category is more of a gradual attenuation than the abrupt decay patterns introduced in §2.3. Whereas the abrupt decay decays all the way immediately outside the trigram window, gradual decay exhibits significant differences among beyond transvocalic co-occurrences.

2.6 Selective decay (blocking)

The last type of unbounded interaction can be characterized as including blocking segments: cases where a ban on a nonlocal sequence is lifted or attenuated by the presence of specific intervening segments. From the perspective of decay, it can be interpreted as if the restriction selectively decays only with a certain set of intervening segments or as if the decay is sensitive to the identity of interveners. The most well known example of this is the role of *r* in Latin *l* dissimilation. The dissimilation process in Latin can be blocked by an intervening, root-final *r* (Dressler 1971, Steriade 1987, Bennett 2013, Stanton 2016a and many others).

(18) Latin L-dissimilation blocked by a root-final *r* (Bennett 2013)

a. [flor-alis] (*[flo**r**-a**r**is]) 'floral'

b. [later-alis] (*[late**r**-a**r**is]) 'lateral'

As can be seen in (18), the suffix surfaces faithfully as [-alis] regardless of the presence of *l* in the root, because an *r* intervenes between the trigger and the target *l*. It has been traditionally generalized that the failure of dissimilation is attributed to the intervening *r*. It is often the case that a pattern exhibits both distance-based decay and selective decay. Latin is precisely the example of the case; the *l* co-occurrence restriction peters out based on the number of intervening syllables as well as the presence of *r* between two *l*s. Thus, the effect of the intervening *r* can be confounded by the distance effect; one might ask whether the *l* co-occurrence in (18) is due to the intervening syllable *r*, which could have been any syllable, or specifically the *r*? The crucial data needed in order to tease these two apart would be examples with one open intervening syllable that has an onset that is not an *r*, such as *plumalis* 'feathered', *glebalis* 'of clods', and *legalis* 'legal', in which the *l* co-occurrence might have been partially licensed due to the distance. I continue this line of discussion in the learning simulation in Chapter 4.

Selective decay is also easily found in vowel interactions. In Shona verbal roots, mid vowels *e* and *o* cannot be followed by a high vowel *i*: *e...i and *o...i. These illegal vowel sequences can actually be licensed by an intervening low vowel *a*, as in [ʧ**e**j**a**m**i**sa] 'make be twisted' and [pof**o**m**a**ʥ**i**ra] 'blind for'.

2.7 Summary

In this chapter, I introduced the typology of long-distance dependencies in phonology from the perspective of locality. I started the description of long-distance phonological dependencies by introducing the previous established dichotomy between trigram-bound and unbounded patterns in consonantal interaction. There is also a similar dichotomy in vowel interactions; vowel harmony can either be non-iterative or fully iterative. This dichotomy that can be observed in both consonantal and vowel interactions is an important underpinning for the learner that I propose in the dissertation.

I also enriched this dichotomous typology by introducing the notion of decay into it; more specifically, I categorize unboundedness into no decay, distance-based decay, and blocking (selective decay), depending on the existence of decay and the cause of decay. Table 2−Table 3 are the summary of the typology. I present the summary in two separate tables due to the limited space.

		Assimilation	Dissimilation
Unbounded	No decay	Berber	*Quechua*
	Gradual decay	Navajo	Latin, *Arabic*
	Blocking	Kinyarwanda	Latin
Trigram-bound		Lamba	Yimas, *Korean*
Adjacent		English	Ainu

Table 2. Consonant typology: alternation *phonotactics*[1]

[1] *Berber*: Hansson (2010b); *Navajo*: Martin (2005); *Kinyarwanda*: Bennett (2013); *Lamba*: Odden (1994); *Quechua*: Gouskova and Gallagher (2020); *Latin*: Bennett (2013), Zymet (2014); *Arabic*: Frisch et al. (2004); *Yimas*: Suzuki (1998), *Korean*: Kim (1985), Ito (2007)

		Assimilation	Dissimilation
Unbounded	No decay	Shona	Malagasy
	Gradual decay	Hungarian	
	Blocking	Shona	
Trigram-bound		Kazakh	Kera, Hebrew
Adjacent		Yoruba	Arusa

Table 3. Vowel Typology: alternation *phonotactics*[2]

It can be seen from the above tables that consonantal interactions and vowel interactions resemble each other in the sense that the dichotomy is robust. Moreover, all sub-categories of unbounded (no decay, gradual decay, and blocker) are attested in both cases. As I will show in the next chapter, this mirroring typology allows my learner to handle consonant and vowel dependencies with a unified approach.

[2] *Shona*: Beckman (1997), *Hungarian*: Zymet (2014), *Kazakh*: McCollum and Kavitskaya (2018), *Yoruba*: Welmers (1973), *Malagasy*: Zymet (2014), *Kera*: Suzuki (1998), *Arusa*: Suzuki (1998)

CHAPTER 3

RESTRICTIVE TIER LEARNER

In the previous chapter, I gave an overview of the typology of long-distance dependencies in phonology from the perspective of locality. I confirmed that consonant and vowel interactions have a mirrored typology. Building on this observation, I propose a phonotactic learner that learns such phonological long-distance dependencies in this chapter: the Restrictive Tier Learner.

In §3.1, I start by explaining in detail the Inductive Projection Learner (Gouskova and Gallagher 2020), which is the base learning algorithm of the learner proposed in this dissertation. In §3.2, I examine what predictions the Inductive Projection Learner makes for each type of interaction laid out in Chapter 2. In §3.3, I characterize the main contribution of my learner: adding the evaluation phase to the existing Inductive Projection Learner (Gouskova and Gallagher 2020). I also introduce the architecture of my algorithm and explain how tiers that are cued by placeholder trigrams are further evaluated in terms of their necessity and accuracy. In §3.4−§3.5, I use natural language examples to demonstrate how adding an evaluation step can benefit learning bounded patterns, unbounded no decay patterns, and selective decay patterns. More specifically, my version of the learner successfully prevents a tier projection for a bounded pattern while allowing for a tier projection for unbounded patterns, and also discovering blockers if there are any. Before wrapping up the chapter, I introduce previous approaches to capturing long-distance phonological dependencies in §3.6. I conclude the chapter in §3.7.

3.1 Inductive Projection Learner (Gouskova and Gallagher 2020)

The Gouskova and Gallagher (2020) learner first induces a baseline (tier-free) grammar using the Hayes and Wilson phonotactic learner (Hayes and Wilson 2008), in order to find evidence that tiers may be needed. The Hayes and Wilson phonotactic learner induces a set of constraints from the learning data by searching through a space of possible constraints. Based on the phonological feature set defined by the analyst, the learner first constructs a list of all natural classes and also an exhaustive list of all possible n-gram constraints (only up a certain n, and $n = 3$ for the Inductive Projection Learner). The learner then combines the language's segments randomly to identify unattested or underattested n-grams in the learning data. The learner induces constraints against these unattested or underattested structures and also weights the constraints, using the principle of maximum likelihood: maximize the probability of the observed patterns in the language.

Regarding the constraint selection criterion, the Gouskova and Gallagher (2020) learner uses *gain* instead of O/E as in Hayes and Wilson (2008). Gain measures how useful adding a certain constraint would be without making any change to the current grammar when accounting for the learning data. Since the learner uses the Hayes and Wilson algorithm as its base algorithm, it adheres to the heuristics of the Hayes and Wilson algorithm in discovering constraints, showing a preference for constraints that are shorter (bigrams over trigrams) and also ones that mention larger natural classes. Thus, in each iteration of a constraint search, while following the order of its preferences, the learner picks out a constraint with the highest gain and also weighs it. Everytime a new constraint is added to the grammar, the learner reweighs all the previous constraints accordingly. There is a threshold for gain that has to be specified by the analyst. When the gain value of the next constraint does not exceed the specified threshold, learning is halted.

Another parameter that can affect learning is *gamma*, which is related to the tolerance of the learner toward exceptions. Higher gamma leads to discovering more exceptionless (high-weighted) constraints whereas lower gamma favors constraints that are violated more often in the data (low-weighted). Relatedly, it is often the case that gamma also affects the specificity or generality of the constraint that the learner discovers; higher gamma leads to finding more specific constraints, which have fewer exceptions and therefore have higher weights. Real-life examples of these cases will be discussed in Chapter 4.

The two parameters, gain and gamma, determine what constraints are discovered in learning simulations and have tremendous influence on the resulting grammar. As Gallagher (2020) notes, it is currently unknown which parameter setting can best mimic human performance in phonotactic learning. Thus, following the strategy that Gallagher (2020) uses, I tried many combinations of those two parameters in the simulations reported in the dissertation. More specifically, I started with gain of 100 and gamma of 0. I increased or decreased gain with an interval of 5 and stopped when the desired constraint was found. After the gain was set tentatively, I raised gamma with an interval of 1 and stopped when the desired constraint could not be found anymore. After gamma was set, I tried the tentative gain value with the new gamma and tweaked them around until the desired constraint could be found stably.

After the baseline grammar search is done, the Inductive Projection Learner goes through the constraints produced by the Hayes and Wilson learner and looks for evidence that a tier projection is needed. The intuition behind this procedure is as follows. The Hayes and Wilson learner augments natural classes by a [word boundary] feature. Word edges are specified as [+word boundary] and all consonants and vowels as [−word boundary] ([−wb] or [ ] henceforth). Gouskova and Gallagher (2020) refer to [−wb] as a *placeholder class* and this is the largest natural class in any language since it includes all consonants and vowels. In a hypothetical language where two *b*s never co-occur across a vowel, phonologists would suggest a constraint such as *bVb. However, the Hayes and Wilson algorithm's preference for a larger natural class will likely to generalize this pattern as *b[ ]b, if one of the following conditions is met. First, if the language has a simple CV structure, *bVb will be easily generalized to *b[ ]b since since neither a vowel nor a consonant would occur between the two *b*s. Secondly, even if the language allows codas and CCC clusters, inducing a more generalized constraint is still possible if these CCC structures are rare enough in the data, which is universally true in languages, and therefore the learner does not see too many bCb sequences. Inducing *b[ ]b rather than *bVb might be not as smooth in the second case but it is still likely since no vowels occur and no consonants occur *enough* between the two *b*s. This kind of trigram constraints with a placeholder class in the medial position, which Gouskova and Gallagher (2020) call a *placeholder trigram*, is a crucial cue to the learner that a projection is needed since it indicates that the segments on either side interact non-locally regardless of the identity of the medial segment.

If placeholder trigrams are included in the baseline grammar, the model creates the tiers based on them. More specifically, the model will look for the smallest natural class that includes both X and Y, which is often either X or Y itself, otherwise a superset of both, and project it. After projecting tiers from the baseline placeholder trigrams, the learner begins discovering constraints anew, looking in turn at both the newly projected tiers and the default tier, where every segment is visible. This process is called *final grammar search* and the resulting grammar is the final grammar. If there is no baseline placeholder trigram or no smallest superset natural class of the either side of the placeholder trigram, the learner does not project tiers and there is no need for final grammar search, either. In this case, learning ends in the baseline grammar, which will be considered the final grammar.

3.2 Predictions of the Inductive Projection Learner

In this section, I examine whether the recent version of the Inductive Projection Learner (Gouskova and Gallagher 2020) can or cannot predict each type laid out in Chapter 2. Using a hypothetical *b* co-occurrence restriction as an example of a consonantal interaction, I show the following; first, I show that the Inductive Projection Learner fails to distinguish transvocalic patterns from unbounded ones and executes a tier projection for both cases. Second, I show that while the Inductive Projection Learner can properly predict the three unbounded types through the weight combination of a baseline trigram constraint and a tier-based bigram constraint, the learner cannot predict *unbounded selective decay*.

The tableau in (19) has five hypothetical forms that represent each of the *b* co-occurrence conditions. The form *basis* includes a single *b* and has no *b* co-occurrence in it. The next three forms, *babis*, *basib*, and *basisib*, all have a co-occurrence of two *b*s with the monotonically-increasing number of intervening consonants. The form *bakib* has a co-occurrence of two *b*s with a consonant *k* in between, which I will use later as an example of blocker.

(19) The violation profile of *b* co-occurrences

	baseline *b[ ]b	*b* tier *[−wb][−wb]
basis	0	0
ba**b**is	-1	-1
basi**b**	0	-1
basisi**b**	0	-1
baki**b**	0	-1

The top row of the tableau shows the tiers on which the Inductive Projection Learner discovers constraints in the final grammar search. On the baseline tier, where every segment is projected, the local trigram constraint *b[ ]b can be learned.

This constraint is violated only by the form *babis* where only one segment intervenes between the two *b*s. The other forms with a *b* co-occurrence do not violate this constraint because they all have more than one segment between the two *b*s. On a tier where only *b*s are visible, the tier-based bigram constraint *[−wb][−wb] can be learned. This constraint is violated by any *b* co-occurrence regardless of the number and the identity of interveners. The four forms with a *b* co-occurrence all equally violate this constraint once.

This toy grammar can yield four different scenarios depending on the combination of the weights of these two constraints, more specifically, whether each constraint has a positive weight or zero. The tableaux in (20) show the harmony scores of the forms in each possible scenario. As mentioned above, what is decisive is whether the weight is zero or a positive value, but two is arbitrarily chosen to represent any positive weight for an illustrative purpose.

(20) Possible scenarios of *b[ ]b on the default tier and *[−wb][−wb] on the b tier

a. *No Restriction*

	*b[ ]b $w=0$	*[ ][ ] $w=0$	$\mathscr{H}$
basis	0	0	0
ba**b**is	-1	-1	0
basi**b**	0	-1	0
basisi**b**	0	-1	0
baki**b**	0	-1	0

b. *Transvocalic*

	*b[ ]b $w=2$	*[ ][ ] $w=0$	$\mathscr{H}$
basis	0	0	0
ba**b**is	-1	-1	-2
basi**b**	0	-1	0
basisi**b**	0	-1	0
baki**b**	0	-1	0

c. *Unbounded No Decay*

	*b[ ]b $w=0$	*[ ][ ] $w=2$	$\mathscr{H}$
basis	0	0	0
ba**b**is	-1	-1	-2
basi**b**	0	-1	-2
basisi**b**	0	-1	-2
baki**b**	0	-1	-2

d. *Unbounded Bi-level Decay*

	*b[ ]b $w=2$	*[ ][ ] $w=2$	$\mathscr{H}$
basis	0	0	0
ba**b**is	-1	-1	-4
basi**b**	0	-1	-2
basisi**b**	0	-1	-2
baki**b**	0	-1	-2

In cases where both constraints have a weight of zero (20a), which I call *No Restriction*, there is simply no restriction regarding *b* co-occurrences; *b*s freely co-occur at any distance and with any interveners. All forms are equally grammatical, with a harmony of zero.

The Inductive Projection Learner can predict the pattern of *Transvocalic* (or transconsonantal for vowel interactions) if the local trigram *b[ ]b has a positive weight and the tier-based bigram has no weight. As shown in (20b), this weight combination yields a pattern where ***b**a**b**is* is ungrammatical whereas the other forms with a *b* co-occurrence, ***b**asi**b***, ***b**asisi**b*** and ***b**aki**b***, are as grammatical as the form without any *b* co-occurrence, ***b**asis*. While this scenario precisely mirrors the bounded transvocalic pattern of McMullin (2016), it can also be seen as an abrupt decay pattern. Since the transvocalic restriction immediately decays all the way and the *b* co-occurrences with extra syllables are no longer underattested at all, the tier-based constraint *bb does not need to be weighted positively. Since there is no need for a tier-based constraint *bb, there is also no need for a tier; a local trigram *b[ ]b is sufficient to capture this pattern. However, the current version of Inductive Projection Learner will always project a *b* tier in this case, because the Inductive Projection Learner will project a tier directly from the baseline placeholder trigram *b[ ]b.

In cases where only the tier-based bigram constraint *[−wb][−wb] has a positive weight, which I call *Unbounded No Decay*, a restriction on a *b* co-occurrence is at work unboundedly, at the same strength without decaying. Looking at the harmony scores in (20c), only ***b**asis* is grammatical with a harmony of zero whereas the other forms are all equally bad with the same negative harmony scores.

If both constraints have a positive weight, the strong restriction holds over one segment and the weaker version of the same restriction holds at a distance that is farther than a segment away, and unboundedly from there without decaying. Looking at the harmony scores in tableau (20d), a transvocalic *b* co-occurrence has the lowest harmony score for violating both constraints. The other forms, ***b**asi**b***, ***b**asisi**b*** and ***b**aki**b***, that have *b* co-occurrences with extra syllables in between, only violate the tier-based bigram and therefore end up being more wellformed than the transvocalic *b* co-occurrence (***b**a**b**is*) yet still more illformed than the form without any *b* co-occurrence (***b**asis*). While it is unclear at the moment whether this pattern has been attested in natural languages, this prediction contrasts with and thus cannot properly capture the already attested distance-based decay pattern, which is more of a gradual attenuation where there are significant differences among beyond transvocalic co-occurrences (Hungarian, Hayes and Londe 2006; Malagasy and Latin, Zymet

2014). For instance, assuming that our hypothetical *b* co-occurrence restriction also follows a gradual decay pattern and the restriction significantly weakens as the number of interveners monotonically increases, the decay pattern of the *b* co-occurrence restriction is only partially captured by this grammar. Under this scenario, the binary distinction between the transvocalic and the beyond transvocalic co-occurrences can be made. However, the differences among beyond transvocalic co-occurrences cannot be captured, because all of these are only penalized by *[−wb][−wb] equally regardless of the number of interveners since tier-based constraints are blind to the baseline information. Thus, the significantly higher wellformedness of ***b**asisi**b***, compared to that of ***b**asi**b*** or ***b**aki**b***, will not be captured. I call this pattern *Unbounded Bi-level Decay* because only the two shades of the *b* co-occurrence restriction, stronger over a single segment and equally strong over more segments, can be reflected in the grammar. Although it is incapable of capturing every shade of distance decay, there are cases where this grammar can be sufficient. If a language has an upper limit on word length, or even if a language allows longer words, if the majority of words in a language is only of a certain length so that it only exhibits a two-level decay, this grammar can be sufficient. Put differently, if the language lacks or has only a small portion of forms like ***b**asisi**b***, the binary distinction between transvocalic and beyond transvocalic would be adequate.

The above descriptions on the tableaux in (20) are summarized in Table 4.

	w(*b[ ]b)	w(*[−wb][−wb])	Scenario
a)	0	0	*No Restriction*
b)	> 0	0	*Transvocalic*
c)	0	> 0	*Unbounded No Decay*
d)	> 0	> 0	*Unbounded Bi-level Decay*

Table 4. Weight combination and the resulting grammar

Notice that there is no weight combination of the given constraint set (*b[]b on the default tier and *[−wb][−wb] on the *b* tier) that can generate the pattern of *selective decay*, in which otherwise illegal *b* co-occurrences are licensed with an intervener *k*. This is because the *b* tier given here is not sufficient to capture patterns with blockers. Since the occurrence of a sequence *bb* is dependent on the existence of *k* in these cases, the licitness of *bb* should be assessed on a tier where both *b* and *k* are visible (McMullin 2016). Take the minimal pair **basib* and *bakib* from the example in (21).

(21) Tier-based representations of **basib* and *bakib*

{b, k} tier	*b b	b k b
{b} tier	*b b	b b
Baseline (all segs)	*basib	bakib

On a tier on which only *b*s are visible, the two forms above have the identical representation (*bb*) and the tier-based constraint **bb* can incorrectly rule out the grammatical form *bakib*. On a tier on which both *b* and the blocker *k* are projected, the two forms have different representations: *bb* and *bkb*. Now, the tier-based constraint **bb* can correctly capture the *selective decay* pattern by prohibiting **basib* while permitting *bakib*. The current version of the Inductive Projection Learner projects a tier directly based on a placeholder trigram (e.g., *b* tier from *b[]b), which is unable to contain information about blockers. The Inductive Projection Learner therefore cannot project blocking segments along with the interacting segments although it is a necessary thing to do in order to capture blocking patterns.

3.3 Adding evaluation

My learner uses the Inductive Projection Learner by Gouskova and Gallagher (2020) as a base algorithm, which in turn was based on the MaxEnt phonotactic learner by Hayes and Wilson (2008). Hayes and Wilson's phonotactic learner identifies unattested or underattested local n-grams in the learning data, up to a certain n that can be user-specified between one and four. In the Hayes and Wilson learner Figure 4(a), tiers need to be manually provided by the analyst in order to capture long-distance dependencies. To this algorithm, which Gouskova and Gallagher (2020) call *baseline grammar search*, Gouskova and Gallagher (2020) add a procedure in which tiers can be automatically discovered, shown in Figure 4(b). The Gouskova and Gallagher (2020) learner has two main components; first, the model goes through the list of constraints produced by the Hayes and Wilson learner and checks if there is evidence that a projection may be needed. Second, the model projects tiers based on the output of the baseline grammar search and builds a final grammar anew by cycling through the default tier and the projected ones, which Gouskova and Gallagher (2020) call *final grammar search*.

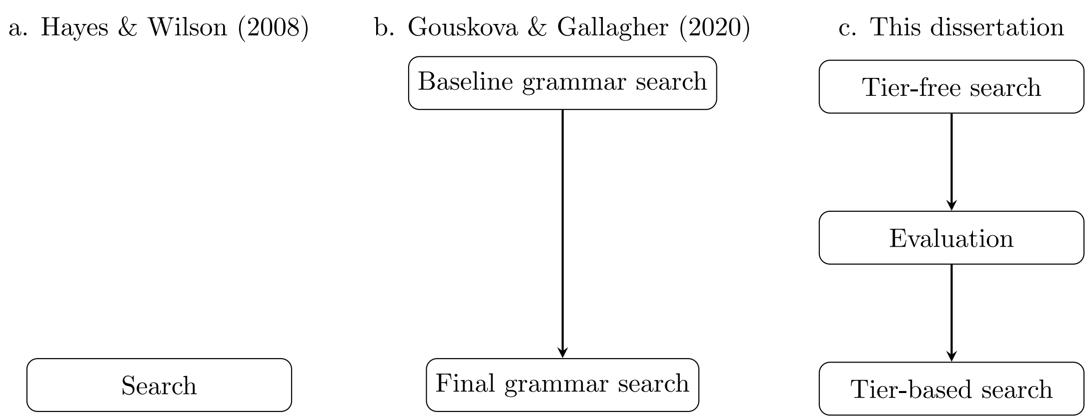

Figure 4. The overview structure the learner, repeated from Figure 1

As intuited and demonstrated by Gouskova and Gallagher (2020), baseline trigrams with a placeholder in the middle (*placeholder trigram*; for example, *X[]Y meaning no co-occurrence of X and Y over any segment) implies that the two natural classes X and Y interact nonlocally regardless of the identity of the middle segment, and thus can provide evidence that learning may benefit from having a separate tier on which only those non-locally interacting natural classes (e.g., X and Y) are visible. According to Gouskova and Gallagher (2020), existing nonlocal restrictions can be learned on those tiers that are automatically induced from placeholder trigrams. I propose in the dissertation that although a baseline placeholder trigram can hint at a projection of a tier, the suggested tier should be further validated in two different ways, before actually being projected in the final grammar search. First, the necessity of a tier projection itself should be examined; the existence of a placeholder trigram does not always mean that having that tier will be beneficial, because the pattern might be simply trigram-bound and not generalize to farther distances. Secondly, the segments to be projected on that tier should also be further examined because what is suggested by the baseline trigram might not be sufficient. That is, the hinted tier might exclude necessary segments, such as blockers.

Thus, I add an extra step, called evaluation, between the baseline grammar search and final grammar search, as illustrated in Figure 4(c). Just as in the Inductive Projection Learner (Gouskova and Gallagher 2020), after the baseline grammar search, which I call *tier-free search*, my learner goes through the constraints that are discovered, checking whether there is any placeholder trigram in the baseline grammar. Unlike the learner of Gouskova and Gallagher where the projection of a tier is automatic from each trigram, my learner goes through *evaluation* before proceeding to a tier projection. Below is an overview of the structure of *evaluation*.

(22) Evaluation overview

```
    Input: a set of baseline placeholder trigrams B
    Output: a set of tiers T
1   Initialize T as empty;
    foreach trigram *X[]Y in B do
3       if IsUnbounded(*X...Y) :
4           a tier t = DetectBlockers(*X...Y)
5           add t to T
        else:
7
    end
```

A brief overview of *evaluation* is presented in (22). Again, as was illustrated in Figure 4(c), evaluation is placed between the two grammar searches. After the tier-free search, the learner selects placeholder trigrams from the learned grammar and feeds them into evaluation. The set of baseline placeholder trigrams is therefore the input to evaluation. Initially, the set of tiers for projection is empty, shown in line `[1]` of (22). With the input placeholder trigram constraints, the learner first determines whether a tier projection is necessary, meaning that the restrictions captured as these trigrams actually generalize beyond the trigram window. If it is determined that the restriction is unbounded, as shown in `[3]`, my learner discovers blockers that need to be on the tier along with the interacting natural classes, shown in `[4]`. Then, as described in `[5]`, the learner projects the tier that includes X, Y, and blockers; if the restriction has blockers, the learner projects a tier based on the interacting natural classes and the discovered blockers. If the learner finds that the restriction is unbounded yet has no blockers, the learner projects a tier based only on the interacting natural classes because blocker will be defined as empty. Finally, if the learner finds that the restriction does not generalize beyond the trigram window, as in [7], the learner does not project a tier based on this trigram. After this process, the learner ends with a set of tiers T, which are necessary and accurate. In the tier-based search, my learner begins discovering new constraints, looking in turn at both the newly projected tiers and the default tier. The learner is available online at `github.com/seoyoungkimm/inductive_projection_learner`.

3.4 Evaluating necessity: is the pattern unbounded?

As the first step of evaluation, the learner determines whether the tier is necessary to capture the pattern observed in the dataset. This is related to the question of whether the pattern captured as a trigram expands unboundedly or not. The tier should be projected only if it does expand, and should not be projected if the pattern is only visible within a trigram window and does not further expand because having a local trigram constraint will be sufficient in the latter case. As I will demonstrate throughout this chapter and Chapter 4, projecting unnecessary tiers can hinder learning. More specifically, it draws the learner's attention to look at unimportant tiers and find trivial constraints on it, which eventually leads to not being able to find actually necessary trigrams on the default tier since the learner is biased towards finding bigrams before trigrams. In addition, too many tiers often prevent the learner from successfully running as they contribute to running out of memory and crashing.

I will first demonstrate how evaluating the necessity of tiers work by using Lamba as an illustrative case. Lamba has a transvocalic nasal harmony in the form of suffixal alternation (Odden 1994); suffixal *l* harmonizes with the nasal *n* in the root, as in [u:**m**-i**n**e] (*[u:m-ile]) 'dried-PERF', compared to [pat-i**l**e] (*[pat-ine]) 'scolded-PERF'. The dependency does not hold at greater distances, as in [**m**as-i**l**e] (*[**m**as-i**n**e]) 'plastered-PERF'; the suffix alternation is not triggered by a nasal segment that is outside the trigram window. This harmony pattern can be easily described by a trigram constraint *[+nasal][ ][+lateral] since the harmony is restricted to transvocalic contexts. Put differently, capturing this pattern does not require a tier-based version of the same constraint (e.g., *[+nasal]...[+lateral]), since it does not hold at arbitrary distances.

According to the typology of non-local interactions in phonology, a restriction captured as a trigram can be considered unbounded if and only if it holds across another segment of the same class. As McMullin (2016) argues, a consonantal restriction observed within a trigram window can be considered unbounded if and only if it holds across another consonant. Similarly to the consonantal restrictions, McCollum (2019) points out that there are only two types of attested vowel harmony patterns in terms of their iterativity; vowel harmony can be either non-iterative where only a single vowel is harmonized within a given domain or fully iterative throughout the domain. He further adds that there is very little evidence for a pattern where n ($n > 1$) vowels may undergo the change within a given domain. Therefore, it can be summarized that consonantal harmony or OCP observed at the trigram level is *guaranteed* to generalize unboundedly, if it holds over another consonant. Similarly, vowel harmony that is observable as a trigram is *guaranteed* to be fully iterative, if it holds over another vowel.

The necessity of a candidate tier is determined by checking whether the importance of the baseline placeholder trigram constraint holds on an intermediate tier which will be projected temporarily for the evaluation purpose. The intermediate tier could be either consonantal or vocalic, depending on the type of interaction that is observed in the trigram; if the interacting segments are consonants, as in *[+nasal][ ][+lateral], a consonantal tier, on which all the consonants and only those are visible, will be temporarily projected. Similarly, if what is captured in the placeholder trigram is a vowel interaction, as in *[+round][ ][−round], a vowel tier, where all the vowels and only those are visible, will be temporarily projected.

In order to determine whether the pattern holds over another segment, discovered placeholder trigrams on the intermediate tier are reweighted, along with the entire baseline grammar on the default tier. One thing to note is that the placeholder [ ] of the tier-based trigram only refers to either any consonant or a vowel, depending on the type of the intermediate tier on which the trigram is reweighted. For example, *[+nasal][ ][+lateral] is equivalent to *[+nasal][+consonantal][+lateral] on the consonantal tier. Crucially, since the same placeholder trigram is simultaneously reweighted on the two different levels, on the temporary evaluation tier and on the default baseline tier, this constraint's weight on the temporary tier will indicate the unboundedness of this restriction. The two tableaux below demonstrate how the unboundedness of a certain restriction is evaluated.

(23) The weights of the trigram on the C tier indicates the unboundedness

a. Trigram-bound

	default tier *n[ ]l $w = 2$	C tier *n[ ]l $w = 0$	$\mathscr{H}$
papile	0	0	0
*panile	-1	0	-2
napile	0	-1	0

b. Unbounded

	default tier *n[ ]l $w = 2$	C tier *n[ ]l $w = 2$	$\mathscr{H}$
papile	0	0	0
*panile	-1	0	-2
*napile	0	-1	-2

The tableaux in (23) represent two possible languages with a simplified phoneme inventory [p n l a i e] that have hypothetical co-occurrence restrictions on nasal and lateral pairs. The language shown in (a) resembles Lamba in which a lateral cannot occur after a nasal within a trigram window; *[panile] is not allowed and always mapped to [panine], confirming to the trigram-bound nasal harmony, whereas [napile] is possible because the stem [n] is farther away from the suffixal [l]. In language (b), a lateral cannot occur if it is preceded by a nasal anywhere in the word. If the co-occurrence restriction is only bounded to a trigram window as in (a), the local trigram *n[ ]l will get a positive weight while *n[ ]l on the consonantal tier will not, since the learner will see numerous datapoints like [napile], which suggests that *n...l does not hold outside the trigram window. Notice that *[panile] is not penalized by *n[ ]l on the consonantal tier because there is no consonant intervening between [n] and [l]. Conversely, if the co-occurrence restriction holds at arbitrary distances as in (b), the weight of *n[ ]l on the consonantal tier will be positive in order to capture the underattestation of words like *[napile]. In both languages (a) and (b), the the weight of *n[ ]l on the default tier remains positive. For the case demonstrated in (a), the local trigram constraint *n[ ]l has to be in charge of capturing transvocalic underattestation, such as in *[panile]. In the case of (b), since *n[ ]l on the consonantal only captures the datapoints C...C...C, the local trigram constraint still needs to do the job of capturing transvocalic datapoints.[1] Hence, the weight of the trigram on the evaluation tier can indicate the unboundedness of the pattern; if the weight of this constraint is 0, it means that such sequences are not underrepresented over another segment; the restriction is merely trigram-bound. On the other hand, if the placeholder trigram receives a positive weight on the temporary tier, it means that such sequences are still underrepresented over another segment, suggesting that the restriction generalizes unboundedly.

The algorithm for determining the necessity of suggested tiers is shown below: (24) is the overview of evaluation, repeated from (22), and (25) provides the architecture of the function `IsUnbounded`, which is used to determine the unboundedness of each restriction learned in the tier-free search.

---

[1]The relative strength between the two constraints shown in (23b) hints at the presence of distance-based decay. If the restriction gradually peters out with the increasing number of interveners, the local trigram *n[ ]l will be weighted higher than the tier-based *n[ ]l in order to capture how the same restriction is enforced more strongly at the higher proximity. If the restriction does not exhibit such distance-based decay, the two constraints will be weighted similarly. The other scenario, in which the local trigram receives a lower weight than the tier-based trigram, predicts an unnatural language where a restriction is enforced more strongly outside the trigram window; it is typologically unattested. A full illustration on the decay and non-decay patterns predictable by the learner is available in (20).

(24) Evaluation, repeated from (22)

```
Input: a set of baseline placeholder trigrams B
Output: a set of tiers T
Initialize T as empty;
foreach trigram *X[]Y in B do
    if IsUnbounded(*X...Y) :
        a tier t = DetectBlockers(*X...Y)
        add t to T
    else:

end
```

(25) The function `IsUnbounded()`

```
def IsUnbounded(*X...Y):
    if X in C and Y in C :
        evaluation tier = consonantal
    elif X in V and Y in V :
        evaluation tier = vocalic
    else:
        continue
    Re-weight *X[]Y on the evaluation tier
    if w(*X[]Y on the evaluation tier) > 0 :
        Unbounded = True
    else:
        Unbounded = False
    return Unbounded
```

As can be seen in the lines `[2]-[6]` of (25), for each baseline trigram constraint, the learner first determines which tier it should be reweighted on. The evaluation tier is chosen based on which natural class, either consonants or vowels, the interacting natural classes both belong to; for example, as seen in the line `[2]`, if both X and Y belong to consonants, the evaluation tier is a consonantal tier. Similarly, the vocalic tier will be chosen as the evaluation tier if both X and Y belong to vowels, shown in `[4]`. This means that reweighting of a trigram will happen only if the two natural classes X and Y both belong to either consonants or vowels. If there is a placeholder trigram found in the tier-free search, one of whose interacting classes is a subset of consonants and the other is a subset of vowels, my learner does not carry the trigram beyond the tier-free search, as shown in the line `[6]`. Put differently, the learner will not evaluate such trigrams for a tier projection. Discussions on the interactions between consonants and vowels can be found in §5.3.

Shown in the lines `[8]-[12]` of (25) is where the unboundedness of a certain pattern is determined. As illustrated with the hypothetical examples above, the weight of the trigram can indicate the unboundedness of the pattern represented in the trigram; positive weight indicates that the pattern is unbounded `[9]` whereas zero weight indicates that the pattern is trigram-bound `[12]`.

In the rest of this subsection, I demonstrate how my learner determines the necessity of tier projection by using examples of unbounded laryngeal restrictions in Quechua and non-iterative round vowel harmony in Kazakh.

(26) Evaluation of the unbounded Quechua laryngeal restriction

Consonantal tier	*k p’	*k m p’
Baseline (all segs)	*kap’i	*kamip’a

Quechua features a three-way distinction for stops and affricates between plain ones ([p t ʧ k q]) and laryngeally marked consonants, which includes ejectives ([p’ t’ ʧ’ k’ q’]) and aspirates ([p^h t^h $ʧ^h$ k^h q^h]). These consonants are under several distributional restrictions. Most widely known are categorical co-occurrence restrictions on laryngeally marked consonants, in which ejectives and aspirates cannot be followed by any of the stops or affricates regardless of the distance between them. Thus, ejectives and aspirates can occur non-initially only if these are preceded by fricatives or sonorants, as in [rit'i] 'snow' and [ʎimp^{h}u] 'clean'. As seen in (26), *[**k**a**p'**i] and *[**k**ami**p'**a], which are hypothetical nonce words of Quchua, are categorically illegal (Gallagher 2016). In the baseline grammar search, these restrictions can be found as *[−sonorant, −continuant][ ][+constricted glottis] and *[−sonorant,−continuant][ ][−continuant,+spread glottis] because the learner will never see obstruent - vowel - ejective and obstruent - vowel - aspirate sequences, finding enough evidence that these pairs are underattested over a vowel in trigrams. Before projecting a tier based on these two trigrams, my learner first determines whether the restriction generalizes unboundedly. As mentioned above, the unboundedness of these restrictions, currently captured at the trigram level, will be assessed by checking whether these restrictions hold across another consonant. This can be done by reweighting these trigrams on a temporarily projected consonantal tier because the placeholder [ ] refers to a consonant on this new level. Since these restrictions are unbounded in Quechua and the learner will never see these pairs over a consonant (e.g., *[kmp'] in *[kamip'a] above), *[−sonorant, −continuant][ ][+constricted glottis] and *[−sonorant,−continuant][ ][−continuant,+spread glottis] will get positive weights on the consonantal tier. After it is confirmed that these restrictions maintain their importance on the new level, implying that the pattern generalizes to arbitrary distances, the learner proceeds to the next step to discover blockers.

Kazakh exhibits non-iterative round vowel armony, in which underlying [+round] vowels trigger harmony exclusively on the following syllable (McCollum 2018; Balakaev et al. 1962). As can be seen in (27), in both forms [mojʊn-də] (*[mojʊn-dʊ]) 'neck-ACC' and [kino-m-ʊz-dəŋ] (*[kino-m-ʊz-dʊŋ]) 'movie-POSS.1-PL-GEN', only a single vowel immediately after a [+round] vowel is harmonized, regardless of the position of the trigger [+round]; the root-internal second syllable vowel is targeted by the first vowel in [mojʊn-də] and the suffix vowel is targeted by the second vowel of the root in [kinomʊzdəŋ].

(27) Evaluation of Kazakh non-iterative vowel harmony

Vocalic tier	o ʊ ə	i o ʊ ə
Baseline (all segs)	mojʊndə	kinomʊzdəŋ

The pattern observed here can be represented as a baseline placeholder trigram *[+round][ ][−round]. If the learner sees enough evidence, for example, underattestation of [+round] - [−round] sequences over a consonant, the constraint can be discovered. However, bounded patterns are hard to be discovered by a phonotactic learner since the learner will see a plethora of counterexamples outside the window in which the restriction is enforced. For example, in order to discover the constraint *[+round][ ][−round] from Kazakh learning data, the learner should see that [+round][ ] sequences are mostly followed by [+round] and not [−round]. In the two forms in (27), [+round][ ] sequences are followed by another [+round] vowel only half the time and [−round] vowel the other half: *o* ([+round]) is followed by *ʊ* ([+round]) twice, and *ʊ* ([+round]) is followed by ə ([−round]) twice. Bounded patterns inherently give rise to an existence of systemic exceptions in learning data and obscure the pattern as an overall phonotactic in a language. Such bounded patterns, however, can become more salient under specific circumstances or made to be more salient by the analyst, which makes these bounded restrictions be discovered more smoothly as a desired trigram in the baseline search. For example, if a language lacks long words and consists mostly of shorter words, the bounded pattern will become more salient without any extra mechanism. If that is not the case, the hyperparameters can be adjusted in such a way that facilitates the discovery of bounded restrictions. Gallagher (2020) assesses the Inductive Projection Learner by testing its ability to induce a strident projection based on the distribution of stridents in various languages. While she reports that the learner is mostly successful, she points out that the parameters of the learner need to be calibrated in a specific way to achieve the desired result. More specifically, gain should be lower to accommodate smaller data sets and gamma should be adjusted accordingly depending on the strength of the pattern (e.g., the number of exceptions). With relatively lenient parameters, bounded patterns can be made more discoverable. I report in §4.2 the successful learning simulation results on Korean, in which dorsals are under a trigram-bound co-occurrence restriction. And a more in-depth discussion about the learnability of trigram-bound patterns is available in §5.1.

Assuming that the trigram *[+round][ ][−round] was successfully discovered in the baseline grammar for Kazakh, it should be determined whether this pattern generalizes unboundedly and a tier projection is necessary, which is not the case here. Similarly to how the unboundedness of Quechua phonotactics was examined, the unboundedness of this pattern can be examined by checking whether this pattern holds across another vowel. Thus, the placeholder trigram constraint is reweighted on a temporarily provided vocalic tier, where all the vowels and only those are visible. The constraint in question, *[+round][ ][−round], can be interpreted as *[+round][+syllabic][−round] because a placeholder refers to vowels on the vocalic tier. Since this vowel harmony is non-iterative, meaning that only a single vowel is harmonized after a trigger vowel, the learner will run into numerous vowel sequences that are [+round][+round][−round], as exemplified well in sequences shown in (27): *oʊə*. This will lead to a very low, even zero, weight of the constraint *[+round][ ][−round] on the vocalic tier. Thus, the learner will be able to confirm that *[+round][ ][−round] does not maintain its importance on the vowel tier. As mentioned above, long-distance vowel interactions can either be non-iterative or fully iterative. Relying on this typological dichotomy, this result implies that the pattern is only restricted to a bounded window. The learner will not project a tier based on this trigram constraint.

As seen so far, whether or not a baseline placeholder trigram constraint is generalizes unboundedly can be determined by re-evaluating its importance on a temporarily provided consonantal or vocalic tier. If the trigram constraint maintains its importance on the temporary tier, a tier projection is executed based on the trigram. If the trigram constraint loses its importance on the intermediate tier, no tier related to this trigram will be projected in the final grammar search.

The evaluation method above relies heavily on the assumption that syllable structures are simple (e.g., CV). If a language exhibits a non-iterative vowel harmony and also allows consonant clusters (e.g., C**V**C.C**V**), my learner will not be able to capture such datapoints. In the evaluation phase, my learner will not proceed to a tier projection since non-iterative vowel harmony does not hold across another vowel. Thus, the final grammar will include only the local trigram constraint but not the tier version of it, mischaracterizing words with consonant clusters. The relevant discussion on languages with consonant and vowel clusters can be found in §5.4.

3.5 Evaluation accuracy: are there blockers?

After the unboundedness of the pattern is confirmed, my learner proceeds to investigate if the pattern is weakened over specific segments. That is, the learner detects blockers in the pattern if there are any. In this case, not only the interacting natural classes but also the blocking segments need to be visible on the tier because the grammar will incorrectly rule out grammatical sequences otherwise. Consider the example of Shona vowel harmony in (28).

(28) Evaluation of Shona vowel harmony

[−low] V tier	e	o	e e	o e	*[e i]	o *[o i]
V tier	e a	o a	e e a	o e a	tʃe a i a	o o a i a
Baseline (all segs)	beka	ʃopa	tʃereŋa	fovedza	tʃejamisa	pofomadzira

Shona includes five distinctive vowels [a e i o u] that are subject to phonotactic restrictions within verbal stems (Beckman 1997; Hayes and Wilson 2008; Gouskova and Gallagher 2020). One of these restrictions defines that mid vowels *e* and *o* cannot be followed by a high vowel *i*, as seen in the first four examples in (28): **e*...*i* and **o*...*i*. This phonotactic restriction can be partially captured by a baseline placeholder trigram *[−high,−low][ ][+high,−back] (*[eo][ ][i]). Before projecting a tier based on this trigram, it should be determined whether this pattern generalizes unboundedly. And if it does, it should be also determined if the restriction can be lifted by certain segments.

In fact, the restrictions **e*...*i* and **o*...*i* can be lifted by an intervening low vowel *a*; as seen in the last two examples of (28), [ʧe̱ja̱m-i̱sa] 'make be twisted' and [pofo̱ma̱ʤ-i̱ra] 'blind for', the high vowel *i* can actually follow an *e* or an *o* if the low vowel *a* occurs in between. Therefore, the vowel sequences **ei* and **oi* are not allowed whereas ✓*eai* and ✓*oai* are allowed. In this case, the low vowel *a* should also be included in the tier because the grammaticality of the subsequences *e*...*i* and *o*...*i* depends on the existence of the vowel *a* in between. Looking again at the last two words in (28), these are mischaracterized on the [−low] tier as ungrammatical because their representations on this tier include illegal substrings: **ei* and **oi*. The tier-based representations of these two words on the full vocalic tier no longer include banned substrings, leading to an accurate evaluation of these forms.

How can a learner detect blockers? Earlier in the chapter, in order to determine the unboundedness of a certain restriction, the placeholder trigram learned in the baseline grammar was reweighted on the consonantal or vocalic tier without further change. The middle placeholder referred to either any consonant or any vowel depending on the type of the intermediate tier, and it was a sufficient level of representation to identify whether the restriction generalizes unboundedly or not. Discovering blockers of a restriction requires more specific representations of the placeholder. For example, in order to correctly capture the blocker of Shona vowel harmony, it is crucial to know that the co-occurrence of a mid vowel and an *i* is contingent on the existence of an intervening vowel *a* specifically, rather than any vowel. Thus, in the phase of detecting blockers, instead of evaluating the importance of the placeholder trigram as it is, what is evaluated is a set of constraints where the middle placeholder is replaced by each of the segments in the language that will be visible on the evaluation tier. If evaluation is executed on the consonantal tier, the placeholder will be replaced by every consonant of the language. Similarly, if the evaluation is executed on the vocalic tier, the placeholder will be replaced by every vowel of the language. For instance, the baseline trigram *[−high,−low][ ][+high,−back] learned in Shona will now be substituted by a set of five different constraints, which is the number of vowels in Shona, seen in Table 5.

Tier	Constraint	Interpretation
vocalic	*[−high,−low][+low][+high,−back]	*[eo][a][i]
vocalic	*[−high,−low][−back,−high][+high,−back]	*[eo][e][i]
vocalic	*[−high,−low][−back,+high][+high,−back]	*[eo][i][i]
vocalic	*[−high,−low][−low,+back,−high][+high,−back]	*[eo][o][i]
vocalic	*[−high,−low][+back,+high][+high,−back]	*[eo][u][i]

Table 5. Shona constraint set for detecting blockers

When evaluating these constraints on the temporary vocalic tier, the weight will be zero if the middle vowel contributes to weakening the restriction. For instance, in Shona, vowel sequences *eai* and *oai* are not underrepresented as the intervening [a] weakens the restriction **ei* and **oi*. Therefore, the first constraint given in Table 5, *[−high,−low][+low][+high,−back], should get a low or zero weight, reflecting the grammaticality of ✓*eai* and ✓*oai*. By contrast, the other four constraints should still get nonzero weights, reflecting the overall underrepresentation of *eVi* and *oVi* where

V is not a low vowel (See more on this point in §4.3). Therefore, by weighting the set of baseline trigram constraints with specific segments as a middle gram on the evaluation tier, blockers of a specific restriction can be discovered. After blockers are identified, not only the interacting natural classes, but also the discovered blockers, should be projected on the tier. For the case of Shona, it leads to projecting the full vowel tier, which includes the interacting mid vowels and the high vowel *i*, as well as the blocker vowel *a*. The algorithm for discovering blockers is presented below. Again, (29) is the overview of evaluation, repeated from (22), and (30) provides the architecture of the function `DetectBlockers`, which discovers blockers for the patterns that passed the unboundedness test in the earlier step of evaluation.

(29) Evaluation

```
Input: a set of baseline placeholder trigrams B
Output: a set of tiers T
Initialize T as empty;
foreach trigram *X[]Y in B do
    if IsUnbounded(*X...Y) :
        a tier t = DetectBlockers(*X...Y)
        add t to T
    else:

end
```

(30) The function `DetectBlockers()`

```
def DetectBlockers(*X...Y):
    initialize tier t as a union of X and Y;
    initialize am empty constraint set as E;
    foreach segment s in the evaluation tier do
        append *XsY to E
    end
    Reweight E on the evaluation tier;
    foreach *XsY in E do
        if w(*XsY) == 0 :
            add s to t
    end
    return t
```

As I will show in the next chapter, this approach of discovering blockers is only partially successful when applied to natural language data; it sometimes picks out more segments as blockers than manually crafted grammars while picking out fewer segments than assumed in other cases. While more test cases in the future will lead to further improvements of this algorithm, I believe it is a good first step to take.

3.6 Previous approaches

So far in this chapter, I introduced the learner of Gouskova and Gallagher (2020) and its predictions, which are building blocks of the learner that I propose in this dissertation. I also illustrated with natural language examples how each step of my learner works. In this section, I summarize previous approaches to capturing long-distance phonological dependencies. At this end of this section, I compare previous models and my model in terms of their capacity.

A large amount of previous work on long-distance phonotactic learning is based on the Subregular Hypothesis, which assumes that all phonological phenomena can be described with a less complex machinery than the full power of regular languages (Heinz 2010). Most of long-distance dependencies in phonology belong to Strictly Local (SL), Tier-based Strictly Local (TSL), or Strictly Piecewise (SP).

SL grammars enforce local dependencies by blocking (accepting) illicit (licit) substrings that are a length of *k*. Interactions between adjacent segments can be represented by SL constraints with varying *k*. For instance, place assimilation in English 'in-' affixation can be represented by a set of strictly local bigrams such as {*np, *nk, *nl, *md, *mk, *ml, *ŋd, *ŋp, *ŋl, and so on}. Trigram-bound patterns in a language with simple syllable structures can be captured as a strictly local trigram; for instance, Lamba nasal harmony could be represented as a set of trigrams {*nal, *nel, *nil, *nol, *nul}.

The class of TSL (Heinz et al. 2011) languages is defined as a class of formal languages wherein a Strictly Local (SL) grammar operates over a tier that is a specific subset of the segments, while ignoring all segments not in that subset. TSL grammars capture non-local dependencies by first projecting a tier and then blocking (accepting) illicit (licit) substrings on the projected tier. Prior to Gouskova and Gallagher (2020), Jardine and McMullin (2017) also offer an algorithm which induces tiers from positive data. Their algorithm, called kTSLIA, can learn the tier and the permitted k-factors (equivalent to positive n-grams) for any k. The major difference between the kTSLIA and the Inductive Projection Learner is that the kTSLIA only has dealt with categorical toy datasets that include segmental representations and has not been tested on noisy natural language data where natural classes can play a huge role. In fact, an earlier version of their learner was introduced in Jardine (2016), which reported that their learner fails to learn any pattern with exceptions, which is common in natural languages. kTSLIA requires perfect data, meaning that the dataset must not have accidental underattestations, in order to give the correct answer; but natural languages do have accidental gaps. Relatedly, Gouskova and Gallagher (2020) noted that kTSLIA will be brittle with stochastic language data while misinterpreting accidental gaps as categorical phonotactic constraints.

Another difference between the Inductive Projection Learner and the kTSLIA is the starting hypothesis about the tier. In the Inductive Projection Learner, there must be a placeholder trigram in order to project a tier and a tier projection is possible from each trigram. Thus, it is both possible that the final grammar might not have any tier at all or have multiple tiers. In kTSLIA, however, it is assumed that there is a single tier where every segment is already projected, to begin with. From there, the algorithm removes each segment from the tier if certain conditions are met. Therefore, it is not possible for a grammar learned from kTSLIA to have multiple tiers.

Whereas SL and TSL languages attend to substrings (e.g., immediate adjacency), SP languages attend to subsequences which are defined over precedence relations among the elements in a string. For the above-mentioned Navajo case, TSL grammar employs a separate representation (tier) where sibilants are visible and defines forbidden substrings on it; e.g., *[α anterior][β anterior]. The tier enables evaluation of non-local sequences in a local manner by only projecting relevant material on it. In comparison, SP grammar evaluates subsequences on the default baseline representation. For example, in Navajo, it is of interest whether [α anterior] *is followed by* [β anterior] somewhere in a string. TSL and SP grammars are representationally very similar in that the distance between the interacting segments or the identity of the interveners are not taken into consideration. However, when it comes to blockers, they make different predictions. Blockers can be accounted for in a TSL grammar if the blockers are visible along with the interacting segments on the same tier but SP grammars can never deal with blockers (Heinz 2010; Dai 2021).

There have been argument for SP grammar over TSL because of the fact that blockers are rare both across and within languages. For example, Bennett (2013) point out that blockers are rare in consonantal harmonies, with one exception of Sanskrit n-retroflexion and even if they seem like they have blockers, it is mostly because of the distance separating the interacting segments, not the nature of the interveners (See also Stanton 2016a about the effect of non-coronal blockers being confounded by other factors such as the distance). However, blockers are widely attested in vowel harmony (Rose and Walker 2011). Thus, if we want to use a unified approach in capturing consonantal interactions and vowel interactions, representing blockers is unavoidable. Another claim against blockers is that, even if a language has blockers, it usually takes up a really small portion of a corpus, which makes the blocking pattern not statistically salient or unimportant (Gouskova and Gallagher 2020). Regardless of its statistical saliency, a grammar must account for blocking patterns if speakers have knowledge about these patterns and productively extend them to noncewords, as shown in Hayes et al. (2009). TSL grammars can provide a computational tool to represent such patterns while SP grammars cannot.

Shown in Table 6 is a comparison between the three tier-inducing algorithms. kTSLIA and the Inductive Projection Learner are on opposite extremes on the scale of restrictiveness; kTSLIA posits a single tier and therefore is not flexible enough to capture complex patterns which require multiple tiers whereas the Inductive Projection Learner projects a tier from every trigram constraint discovered in the first grammar search and therefore is too lenient, resulting in a projection of too many tiers. The learner proposed by this dissertation will improve the restrictiveness of the Inductive Projection Learner by adding an extra step where the necessity of candidate tiers are further evaluated before being projected. The Inductive Projection Learner is not capable of detecting blockers of a given pattern because tiers are projected based on a trigram, which is too short to include evidence of blockers. kTSLIA is able to detect blockers only if the data has no exceptions, which is a condition that can rarely be met by natural languages. The learner suggested by this dissertation is able to detect blockers that are not clear-cut.

	kTSLIA	IPL	This dissertation
Naturalistic data	✗	✓	✓
Blocker detection	✓	✗	✓
Multiple tiers	✗	✓	✓
Analyst input	length of k	feature chart, gamma, gain	

Table 6. Comparison between the three tier-inducing algorithms

The last point of comparison regards input from the analyst. In kTSLIA, the analyst has to specify the k value as a certain constant whereas the constraint lengths are set as 2 and 3 in the other two algorithms. Because kTSLIA deals only with categorical restrictions, there are no hyperparameters that need to be specified for kTSLIA. The other algorithms require that the analyst manually choose gamma and gain; these make tremendous impact on the learning results (Gallagher 2020) as they determine what constraints make it into the final grammar.

3.7 Summary

In this chapter, I propose the main contribution of my learner: adding an extra evaluation step to the existing Inductive Projection Learner (Gouskova and Gallagher 2020). This idea is built upon the typological observation that non-local dependencies that can be seen in consonantal interactions and vowel interactions are either trigram-bound or unbounded. The approach taken in the learner of Gouskova and Gallagher, which is automatically projecting a tier from a trigram, can guarantee a maximally general hypothesis. However, it is not necessarily sufficiently restrictive since it can lead to projecting unnecessary tiers; this calls for an addition of the extra step for further inspecting the validity of the candidate tiers.

In the evaluation phase, the tiers cued by baseline trigrams are therefore examined in terms of their necessity and accuracy. The necessity of the tier is related to the question of whether the restriction that is captured as a trigram actually generalizes outside that window (e.g., the restriction is unbounded). The accuracy of the tier is related to the question of whether the tier includes all the necessary segments that are required to be projected, such as blockers.

A projection of a temporary consonantal or vocalic tier in the evaluation step aids the inspection of the candidate tiers. The necessity of the tier can be examined by reweighting the baseline placeholder trigram itself on the temporary tier. The accuracy of the tier can be examined by a set of trigrams, whose middle placeholder is replaced by every segment that is visible on the evaluation tier. By exploiting a typological observation as a heuristic, the learner can successfully project tiers more restrictively and accurately.

CHAPTER 4

LEARNING

In the previous chapter, I introduced the Restrictive Tier Learner, which is equipped with an extra step in which the candidate tiers are evaluated in terms of their necessity and accuracy. I provided natural language examples to demonstrate how the learner distinguishes between trigram-bound and unbounded patterns in §3.4 and how the learner discovers blockers for unbounded restrictions in §3.5. The learner is available online at `github.com/seoyoungkimm/inductive_projection_learner`.

In this chapter, I present the learning simulation results produced by my proposed algorithm; I will also provide learning results of the Inductive Projection Learner when comparison is necessary. When reporting the simulation results, Gouskova and Gallagher (2020) evaluated the resulting grammar holistically; they generated grammatical and ungrammatical testing items and checked whether their grammar successfully made separations between them, using statistical methods. In my dissertation, I take an alternative approach; I check whether desired constraints are included in the grammar. The patterns that my learner captures differently than the Inductive Projection Learner are trigram-bound and blocking patterns, which usually do not stand out as a salient pattern within a language data. While capturing these might not statistically improve the overall fit of the grammar given a corpus, I believe that the grammar should be still able to represent them and a computational learner should capture them if they are psychologically real for language speakers. For example, if there is a dataset from an experimental study that focuses on the role of blockers, the fit of the grammar is expected to be significantly better. The idealized learning results for each typological pattern is summarized in Table 7.

		Unbounded no decay	Trigram bound	Blocking	Distance decay
		§4.1	§4.2	§4.3	§4.4
Tier-free search		trigram	trigram	trigram	trigram
Unbounded test		pass	fail	pass	pass
Blocker test		{}	N/A	blockers	{}
Tier-based search		bigram	N/A	bigram	bigram, trigram
Resulting grammar	tier	bigram		bigram	bigram
	baseline		trigram		trigram

Table 7. Ideal learning scenarios

In §4.1, I report a case study on the laryngeal co-occurrence restriction in Quechua, which is an example of the *unbounded no decay* type. I show that my algorithm motivates the projection of a relevant tier for such patterns and discovers bigrams on the tier, which is necessary to capture dependencies that hold at arbitrary distances. In §4.2, I report a case study on Korean place OCP as an example of *trigram-bound* pattern. While the Inductive Projection Learner automatically projects an unnecessary tier once a placeholder trigram is discovered in the tier-free search, my learner successfully prevents an unnecessary tier from being projected. I also briefly discuss conditions that need to be met in order for such trigram-bound patterns to be correctly captured by the learner. In §4.3, I report a case study on Shona, with a focus on the role of the low vowel [a] as a blocker in height harmony. The learner makes an accurate prediction about the low vowel [a], reflecting the traditional description of the blocking pattern. The learner also predicts [u] to be a blocker, which has not been previously reported to be one in the literature. I examine the causes of this prediction. In §4.4, I report simulation results on Arabic, which was used as a case of bi-level distance decay. While this is a pattern that the Inductive Projection Learner can already smoothly deal with, I show that the addition of evaluation also does not harm the learning of such patterns. Lastly, I report the Latin case study in §4.5. Latin is a complicated case, as its L-dissimilation process exhibits distance decay and selective decay, while the lexicon also has transvocalic phonotactic restrictions. My learner performs better in capturing the distance decay and the trigram-bound aspects of the Latin data, compared to the Inductive Projection Learner. Discovering blockers, however, is only partially successful; I show the fragility of the current method of discovering blockers. I conclude the chapter in §4.6.

4.1 Unbounded no decay: Quechua

I first demonstrate how the addition of the evaluation phase guarantees that relevant tiers are correctly projected for the unbounded pattern. I use the laryngeal co-occurrence restrictions of Quechua as the test case.

Quechua features a distinction for stops (plosives and affricates) between plain ones [p t ʧ k q] and laryngeally marked stops, which includes ejectives [p' t' ʧ' k' q'] and aspirates [p^h t^h $ʧ^h$ k^h q^h]. As mentioned in Chapter 3, Quechua has several restrictions on the distribution of these cosonants. Most widely known are categorical co-occurrence restrictions on laryngeally marked consonants, in which ejectives and aspirates cannot follow any of the plain stops regardless of the distance between them; ejectives and aspirates can occur non-initially only if these are preceded by fricatives or sonorants, as in [rit'i] 'snow' and [ʎimp^{h}u] 'clean'. Some nonce forms of Quechua are presented in (31), repeated from (26). As can be seen in these forms, *[**k**a**p'**i] and *[**k**ami**p'**a] are both categorically illegal in this language (Gallagher 2016).

(31) Evaluation of the unbounded Quechua laryngeal restriction

stop tier	*k p'	*k p'
Baseline (all segs)	*kap'i	*kamip'a

Categorical and unbounded restrictions like these can be smoothly learned with the learner of Gouskova and Gallagher. Since the pattern is inviolable, meaning that there are no exceptions exhibited in the learning data, the restriction easily stands out as an overall phonotactic pattern of the language. Once the restriction is discovered at the trigram level, it is guaranteed that a tier will be provided in the final grammar search. On the projected tier, the non-local versions of the same restrictions can be learned since the patterns do not decay based on the distance between the interacting segments; the restrictions remain salient in the tier representations. This was proven true in the Quechua case study reported in Gouskova and Gallagher (2020); the Inductive Projection Learner smoothly discovered the laryngeal phonotactics as placeholder trigrams *[−sonorant, −continuant][ ][+constricted glottis] and *[−sonorant, −continuant][ ][−continuant, +spread glottis] in the baseline grammar search. The Inductive Projection Learner then proceeded to an automatic tier projection of [−sonorant, −continuant], which is the smallest superset natural class of the interacting natural classes in both of these constraints; both [+constricted glottis] and [−continuant, +spread glottis] are subsets of [−sonorant, −continuant]. On this tier, the unbounded version of these restrictions can be captured. Gouskova and Gallagher (2020) report their success on Quechua; both *[ ][+spread glottis] and *[ ][+constricted glottis] are learned on the projected tier and highly weighted in the final grammar search, correctly capturing the full range of the restrictions in the language. I will show in this section that a projection of the necessary tier is also guaranteed in my algorithm. More specifically, the evaluation method correctly determines the unboundedness of the Quechua laryngeal restrictions by reweighting the two baseline trigrams on the temporarily projected consonantal tier. My learner exploits non-zero weights of the baseline trigrams as evidence that tier projection is indeed necessary. Finally, my learner also finds non-local versions of the same restrictions on these projected tiers in the final grammar search.

The learning data is identical to the one that Gouskova and Gallagher (2020) use and is available at `github.com/gouskova/inductive_projection_learner`. The data includes 10, 848 phonetically transcribed Quechua words compiled from 31 issues of a Bolivian Quechua newspaper.

In order to replicate the learning results of Gouskova and Gallagher (2020), I set the gain at 150, following their case study on Quechua. Another setting that had to be provided to serve the Quechua case is the optional parameter "inviolable". The underlying MaxEnt algorithm (Hayes and Wilson 2008) is forced to only discover exceptionless generalizations in the learning data with this option. Without this parameter, the learner still finds the same trigram constraints with a very high gamma value around 70, which basically has the same effects as providing the "inviolable" parameter; the high gamma biases away from finding constraints with lower weights, put differently, constraints that are prone to being violated. With gamma values that have been typically used for cases studies in Gallagher (2020), the learner found slightly more general constraints. The discussions about the generality of constraints can be found in §5.2 and I report the model with the "inviolable" setting in this chapter. My learner first discovered the desired trigram constraints, as presented below. The two constraints have high weights, reflecting the fact that these restrictions lack exceptions.

Constraint	Interpretation	Weight
*[−cont, −son][ ][+cg]	*[stop][any seg.][eject.]	14.2
*[−cont, −son][ ][−cont, +sg]	*[stop][any seg.][asp.]	14.2

Table 8. Quechua: tier-free search

My learner then proceeds to evaluation. Since the restrictions captured as placeholder trigrams describe consonantal interactions, it is assumed that a temporary consonantal tier, on which all the consonants and only those are visible, is projected in the evaluation phase. First, in order to confirm the unboundedness of these patterns, the two trigram constraints in Table 8 are reweighted on the consonantal tier, along with the entire baseline grammar on the default tier. The result is presented below.

Tier	Constraint	Interpretation	Weight
default	*[ ]	*[any segment]	0
⋮	⋮	⋮	⋮
default	*[−cont,−son][ ][+cg]	*[stop][any seg.][eject.]	15.3
default	*[−cont,−son][ ][−cont,+sg]	*[stop][any seg.][asp.]	14.8
⋮	⋮	⋮	⋮
consonantal	*[−cont,−son][ ][+cg]	*[stop]...[C]...[eject.]	15.4
consonantal	*[−cont,−son][ ][−cont,+sg]	*[stop]...[C]...[asp.]	14.9

Table 9. Quechua: unboundedness test

The placeholder trigram constraints are highly weighted on the two different levels: default tier and the consonantal tier. Whereas the trigram on the default tier exclusively captures strictly trigram-bound and transvocalic underattestation (e.g., CVC), the trigram on the consonantal tier captures underattestation across another consonant (e.g., C...C...C). Most importantly, the high weights on the consonantal tier imply that these restrictions generalize unboundedly. Based on this result, the learner now proceeds to discovering blockers.

As the second step of evaluation, the learner investigates if the restrictions that passed the unboundedness test can be weakened by any interveners. That is, whether the restrictions have blockers. For the Quechua case, it is not expected that the learned restrictions are weakened by certain intervening consonants; as mentioned above, these restrictions in Quechua are categorical. The middle placeholder of the two trigram constraints in Table 8 was replaced by each of the 34 distinctive Quechua consonants. And these 68 trigrams were reweighted on the consonantal tier. As expected, all the constraints received high weights around 12, reconfirming the exceptionlessness of the restrictions at hand.

Since the learner finds that the trigrams in Table 8 generalize unboundedly and have no blockers, the learner projects a tier based only on the interacting natural classes of these trigram constraints, as opposed to, for example, projecting a tier based on interacting natural classes and discovered blockers. This leads to projecting a tier of [−sonorant, −continuant], on which plain, ejective, and aspirated stops are visible. Finally, in the tier-based search, the learner finds the desired bigram constraints that capture the phonotactic restrictions that hold at any distance. The results are presented below.

Tier	Constraint	Interpretation	Weight
[−sonorant, −continuant]	*[ ][+cg]	*[stop]...[ejective]	16.5
[−sonorant, −continuant]	*[ ][+sg]	*[stop]...[aspirate]	16.2

Table 10. Quechua: final grammar

Unbounded no decay patterns are easy to discover by a computational learner, as they easily stand out as exceptionless generalizations in the learning data. In this case study, my learner did not add anything compared to the learning results of Gouskova and Gallagher; the Inductive Projection Learner already handles these cases very smoothly. However, the addition of evaluation did not prevent learning the correct tier either. Thus, it can be concluded that evaluation is a harmless process to be included.

4.2 Trigram-bound: Korean place OCP

In native Korean words, homorganic consonants tend to be avoided across a vowel. This restriction is most strongly attested with dorsals and most strict in monosyllabic words (Kim 1985; Ito 2007; Kang 2015). For example, there are only three exceptions in the Korean lexicon where a monosyllabic native Korean word has two dorsals co-occurring over a vowel: [k^hoŋ] 'bean', [koŋ] 'ball', and [kuk] 'soup'. Dorsals freely co-occur outside the trigram window, as in [kalki] 'mane', [kicike] 'stretch', and many others. Capturing a pattern like this does not require a tier projection since the co-occurrence restriction is enforced only within a trigram window; a trigram constraint *[+dorsal][ ][+dorsal] will be sufficient. If the Inductive Projection Learner discovers this constraint in the tier-free grammar search, however, it will proceed to an automatic projection of a tier on which dorsal segments are visible: e.g., [+dorsal]. I first show in this section that the Inductive Projection Learner indeed runs into the issue of projecting an unnecessary tier when the pattern is merely trigram-bound. I discuss potential problems that can arise from it. Subsequently, I show that the evaluation phase determines that this restriction is only bounded to trigrams and successfully prevents an unnecessary tier from being projected.

The dataset comes from Park (2020). It includes 1,630 native Korean monomorphemic common nouns that are listed in the corpus *Korean Usage Frequency*, whose token frequency is above 5. These words are listed as their standard pronunciation forms, or as how they are provided in a major Korean dictionary, which reflects Korean phonology. The data is available as a UCLAPL-friendly format at `www.linguist-nayoung.com/data`.

I first investigate if the dataset I am using reflects the description about the trigram-bound co-occurrence restriction on dorsals. I compute the O/E values of [+dorsal]...[+dorsal] pairs with and without intervening consonants in the dataset, as presented in Table 11. As seen in the table on the left (a), two dorsal consonants are underrepresented when there is no consonant in between. Dorsal pairs are not specifically underrepresented with one or more intervening consonants, as can be seen in the table on the right (b). The O/E values in Table 11 support the description that the dorsal co-occurrence restriction is enforced only within a trigram window.

	(kkʰk'ŋ)
(kkʰk'ŋ)	.6

a. 0 consonant away

	(kkʰk'ŋ)
(kkʰk'ŋ)	1.1

b. 1 or more consonant away

Table 11. Korean: O/E values of [+dor]...[+dor] pairs at different distances

As discussed in §3.4 above, learning of a trigram-bound pattern involves inherent difficulties because it does not stand out as an overall phonotactic of the language in the input data due to its nature of being bounded. For example, since the restriction is trigram-bound, longer words will unavoidably include trigrams that serve as counter evidence, such as in [cokak] 'piece', [p'ʌk'uki] 'cuckoo', and many others; the evidence for *[+dorsal][][+dorsal] is only partial in longer words. There is an extra complication for this Korean case, which is that the evidence is also partial across the lexicon. As mentioned above, the restriction is most salient in monosyllabic words, meaning that it is enforced more strongly to the specific sublexicon of monosyllabic words. Gallagher (2020) points out that parameters need to be calibrated to cater to different datasets and patterns. For instance, gain should be lower to accommodate smaller data sets and gamma should be lower if the pattern has more exceptions (e.g., less salient). The Korean dataset is pertinent to both conditions; it is of a fairly small size and the strength of the pattern is weak, as the co-occurrence restriction is lifted outside the trigram domain. Therefore, gain and gamma were both adjusted down to handle the characteristics of the Korean dataset.

First, I ran a simulation on the Inductive Projection Learner with a gain of 15 and a gamma of 0. The learner found *[+dorsal][][+dorsal] ($w = 1$) in the baseline grammar search. And as expected, the discovery of this trigram led to an automatic projection of a [+dorsal] tier in the final grammar search. In the final grammar search, the learner found a number of low-weighted bigrams on the [+dorsal] tier, as shown in Table 12, which mischaracterize the observed dorsal OCP pattern as unbounded. And most importantly, the trigram *[+dorsal][ ][+dorsal], which is actually required to properly capture the trigram-bound nature of this restriction, was not discovered again on the default tier.

Tier	Constraint	Interpretation	Weight
[+dorsal]	*[+laryngeal][ ]	*[k^hk']...[kk^hk'ŋ]	0.3
[+dorsal]	*[+aspirate][ ]	*[k^h]...[kk^hk'ŋ]	1.6
[+dorsal]	*[+sonorant][ ]	*[ŋ]...[kk^hk'ŋ]	0
[+dorsal]	*[−tense][−sonorant]	*[kk^h]...[k^hk']	0.5
[+dorsal]	*[−aspirate −tense][+wb]	*[k]...[word boundary]	0.1

Table 12. Korean: final grammar by the Inductive Projection Learner

A projection of an unnecessary tier can distract the learner. Since the Inductive Projection Learner uses the Hayes and Wilson algorithm as its base algorithm, it adheres to the heuristics of the learner of Hayes and Wilson in discovering constraints, showing a preference for constraints that are shorter (bigrams over trigrams). For example, with the [+dorsal] tier projected in the final grammar search, the learner is biased towards discovering bigrams on the this tier before discovering local trigrams on the default tier. In the tier-based representation on the [+dorsal] tier, the learner can no longer distinguish between transvocalic dorsal pairs (kVk configuration) and dorsal pairs with intervening non-dorsal consonants (e.g., kVtVk) because the learner is blind to the existence of the intervening non-dorsal consonants. The underattestation of dorsal co-occurrence over a vowel contributes to lowering the overall O/E value of dorsal pairs at any distance, which facilitates discovering tier-based bigrams on

[+dorsal], such as the ones that are shown in Table 12. Critically, once these bigrams are discovered, the learner will not discover the local trigram *[+dorsal][ ][+dorsal] on the default tier. Since the tier-based bigrams do the work of capturing the underattestation of any dorsal pairs regardless of the distance, both trigram-bound and unbounded, the learner will not try further to discover *[+dorsal][ ][+dorsal]; for example, *[+dorsal][ ][+dorsal] will not improve the fit of the grammar.

I now report the Korean simulation result produced using my learner and show how my algorithm with an extra evaluation step successfully prevents projecting an unnecessary [+dorsal] tier, unlike the Inductive Projection Learner (Gouskova and Gallagher 2020). First, my learner discovered the local trigram *[+dorsal][ ][+dorsal] (w = 1.1) in the tier-free search with the same set of parameters: gain of 15 and gamma of 0. As the first step of evaluation, the learner reweights *[+dorsal][ ][+dorsal] on the temporarily provided consonantal tier along with the entire tier-free grammar on the default tier. The result is shown below.

Tier	Constraint	Interpretation	Weight
default	*[ ]	*[any segment]	1.2
default	*[+laryngeal]	*[p^{h}p’t^{h}t’c^{h}c’k^{h}k’s’]	1.9
⋮	⋮	⋮	⋮
default	*[+dorsal][ ][+dorsal]	*[kk^{h}k’ŋ][ ][kk^{h}k’ŋ]	0.9
⋮	⋮	⋮	⋮
consonantal	*[+dorsal][ ][+dorsal]	*[kk^{h}k’ŋ]...[C]...[kk^{h}k’ŋ]	0

Table 13. Korean: unboundedness test

Table 13 shows the result of evaluating the necessity of the candidate tier [+dorsal]. As can be seen, the consonantal tier version of the trigram was weighted zero whereas the local trigram received some positive weight. The literal interpretation of this result is that a subsequence of two dorsals is underattested within a trigram window but no longer underattested if a consonant intervenes. The implication that can be drawn from these weights is that the dorsal co-occurrence restriction does not generalize at arbitrary distances, which in turn suggests that a tier projection from this specific trigram is unnecessary. Since the restriction turns out to be only trigram-bound, the learner does not proceed to finding blockers or projecting a tier. In other words, after the boundedness of a restriction is confirmed, learning ends with the tier-free grammar.

The inclusion of the evaluation step can help with determining whether a candidate tier is actually necessary or not. The prevention of projecting an unnecessary tier can improve the learning results, as it does not distract the learner away from finding the local trigram that is actually necessary.

4.3 Unbounded with blockers: Shona

As shown in Chapter 3, if the learner confirms that a restriction captured as a trigram actually generalizes unboundedly, it proceeds to look for blockers before projecting a tier. The projected tier will include blockers, which allows for the final grammar to evaluate strings that include intervening blockers more accurately. In this section, I demonstrate how the evaluation step detects blockers given language data. It is well known that blockers are more common in vowel interactions than in consonantal interactions (Rose and Walker 2011, Bennett 2013). As an example of vowel harmony that has blockers, I use Shona as a representative case; Shona vowel harmony is unbounded, fairly exceptionless, and has blockers. Moreover, Shona was a test case in both of the previous studies that inspired the current work: Hayes and Wilson (2008) and Gouskova and Gallagher (2020). Hayes and Wilson (2008) show that manually providing a vocalic tier is unavoidable in capturing nonlocal phonotactics of Shona vowel harmony. Gouskova and Gallagher (2020) successfully show that decision making by the analyst about the tier projection can be removed in their algorithm because tiers can be automatically induced from local trigrams. In their paper, the Inductive Projection Learner successfully discovers a number of vowel tiers; but these tiers do not include blockers, which leads to mischaracterizing strings with an opaque segment. I aim to show in this section how the addition of an extra evaluation step aids the learner in discovering blockers of the restrictions, which results in producing a more accurate grammar. Similar to the structure of section §4.2 above, I first present the simulation results of the Inductive Projection Learner; the projected tier excludes the blocker of the pattern and the grammar detrimentally penalizes numerous datapoints. Subsequently, I report the simulation result produced by my learner in which the projected tier includes the blocker.

Following Hayes and Wilson (2008) and Gouskova and Gallagher (2020), I also restrict the learning data to Shona verbal stems; the harmony pattern is nearly exceptionless in verbs, and even more exceptionless in verbal stems. The learning data that I obtained from Gouskova and Gallagher includes 4,688 verbal stems. As mentioned in the previous chapter, Shona vowels are subject to various phonotactic restrictions. I provide a brief summary of the patterns that have been reported in the previous literature in (32). The Shona examples provided below are from Hayes and Wilson (2008).

(32) Shona vowel distribution

a. The distribution of [a] is not restricted

b. Mid and high vowels [e o i u] can occur in initial syllables

   i. [beka] ‘belch’

   ii. [gondwa] ‘become replete with water’

   iii. [gwiʃa] ‘take away’

   iv. [huna] ‘search intently’

c. Mid vowels [e] and [o] can occur noninitially if preceded by other mid vowels; but [o] can only strictly follow another [o].

   i. [cherenga] ‘scratch’

   ii. [fovedza] ‘dent’

   iii. [dokonya] ‘be very talkative’

d. The high vowel [i] can occur noninitially only if preceded by non-mid vowels

   i. [kabida] ‘lap (liquid)’

   ii. [bhigidza] ‘hit with thrown object’

   iii. [churidza] ‘plunge’

e. The high vowel [u] can occur noninitially only if preceded by non-mid vowels or [e]: *ei, *oi, *ou, ✓eu

   i. [baduka] ‘split’

   ii. [bikura] ‘snatch and carry away’

   iii. [dhuguka] ‘cook for a long time’

   iv. [chevhura] ‘cut deeply with sharp instrument’

Among these distributional restrictions on vowels given in (32), I focus on the specific restriction (d), in which the mid vowels cannot be followed by the high vowel

[i]. This restriction *[eo]...[i] is known to be weakened by an intervening low vowel [a] (Beckman 1997; Hayes and Wilson 2008); for example, the sequences of **ei* and **oi* are illegal whereas ✓*eai* and ✓*oai* are allowed. Notably, the front and back mid vowels do not behave symmetrically; as described in (32d-e), the mid vowels cannot be followed by the high vowels [i u] in general but the front mid vowel [e] can in fact be followed by [u]. Moreover, briefly described in (32c), mid vowels need to follow another mid vowel in general but [o] needs to specifically follow another [o]. As I will show later in the section, this asymmetrical behavior of mid vowels poses extra complexity in learning the Shona vowel harmony.

I first ran a simulation on the Inductive Projection Learner in order to show how tiers projected in this learner allow these grammatical sequences (✓*eai* and ✓*oai*) to be mischaracterized as illegal. The size and near exceptionlessness of the Shona data allow the use of higher gain in learning simulations. With a gain value of 190 and a gamma of 0, the learner discovered all the restrictions provided above. The learning results are shown in Table 14.

Constraint	Interpretation	Weight
*[+high][ ][−high,−low]	*[high V][ ][mid V]	5.5
*[+low][ ][−high,−low]	*[a][ ][mid V]	4.5
*[−high,−low][ ][+high,−back]	*[mid V][ ][i]	3.9
*[−high,−low,+back][ ][+high,+back]	*[o][ ][u]	3.3

Table 14. Shona: baseline grammar by the Inductive Projection Learner

In the table above, the constraint *[−high,−low][ ][+high,−back] captures the restriction of interest, described in (32d): **ei* and **oi*. The last constraint captures more specific version of the restriction in which [u] cannot follow [o]. The first two constraints describe the restriction given in (32c), in which mid vowels can occur only if these are preceded by another mid vowel. Based on these four trigram constraints with a placeholder in the middle, the Inductive Projection Learner projected three tiers that are summarized in Table 15, on which tier-based local bigrams are found in the final grammar search.

Constraint	Tier projection
*[+high][ ][−high,−low], *[−high,−low][ ][+high,−back]	[−low]
*[+low][ ][−high,−low]	[−high]
*[−high,−low,+back][ ][+high,+back]	[−low, +back]

Table 15. Shona: tier projection by the Inductive Projection learner

Most importantly, the placeholder trigram *[−high,−low][ ][+high,−back], which captures the restriction of **ei* and **oi*, led to projecting a tier excluding the low vowel [a], namely the [−low] tier. In the final grammar search, the Inductive Projection Learner finds a bigram on this tier, as shown in Table 16.

Tier	Constraint	Interpretation	Weight
−low	*[−high][ ][+high,−back]	*[eo]...[i]	3.7

Table 16. Shona: final grammar by the Inductive Projection Learner

This bigram on [−low] can incorrectly rule out the otherwise illegal sequences that are licensed by an intervening [a], such as *eai* and *oai*. As schematized in (33), since the [−low] tier is agnostic to the presence of the low vowel [a], the grammar will penalize words that are actually allowed, such as [ʧ**e**j**a**m**i**sa] and [pof**o**m**a**ʥ**i**ra].

(33) Evaluation of Shona vowel harmony

[−low] tier	*[e i]	o *[o i]
Baseline (all segs)	✓ ʧej**a**misa	✓ pofom**a**ʥira

Now I return to my algorithm and demonstrate how evaluation discovers blockers using the learning data. As mentioned in the previous chapter, my learner finds blockers for the restrictions that are already confirmed to be unbounded. For determining the unboundedness of the pattern, the baseline placeholder trigram was reweighted on the temporary tier without further modification. However, at the second part of evaluation after the unboundedness test, more specific representations of the middle placeholder are required in the baseline trigrams. This is because our interests lie in the identity of specific segments that allow otherwise forbidden structures to occur. For example, in Shona, we need to be able to distinguish between the low vowel [a] and the other vowels, in terms of whether they allow *ei* and *oi* to co-occur across them. Therefore, as mentioned in §3.4, the middle placeholder of the trigram in evaluation is replaced by every segment that will be visible on the intermediate tier. And this set of trigrams, instead of a single placeholder trigram, is reweighted.

In the tier free search, my learner found all the constraints that the Inductive Projection Learner discovered, with the gain of 195, as summarized in Table 17.

Constraint	Interpretation	Weight
*[+high][ ][−high,−low]	*[high V][ ][mid V]	5.5
*[+low][ ][−high,−low]	*[a][ ][mid V]	4.5
*[−high,−low][ ][+high,−back]	*[mid V][ ][i]	3.8
*[−high,−low,−back][ ][+high,+back]	*[o][ ][u]	3.9

Table 17. Shona: baseline grammar

As the first step of evaluation, the trigrams in Table 17 are assessed for their unboundedness. Since the interacting natural classes of these trigrams all belong to vowels, the vocalic tier is provided as the evaluation tier. Importantly, they all gained positive weights on this vocalic tier, correctly reflecting the known properties of Shona vowel harmony, which is that they hold at arbitrary distances within verbal stems. The result of the unboundedness test is summarized in Table 18.

Tier	Constraint	Interpretation	Weight
default	*[ ]	*[any segment]	1.0
⋮	⋮	⋮	⋮
default	*[+high][ ][−high,−low]	*[high V][ ][mid V]	4.8
default	*[+low][ ][−high,−low]	*[a][ ][mid V]	4.0
default	*[−high,−low][ ][+high,−back]	*[mid V][ ][i]	3.5
default	*[−high,−low, +back][ ][+high,+back]	*[o][ ][u]	3.4
⋮	⋮	⋮	⋮
vocalic	*[+high][ ][−high,−low]	*[high V]...[V]...[mid V]	2.6
vocalic	*[+low][ ][−high,−low]	*[a]...[V]...[mid V]	2.9
vocalic	*[−high,−low][ ][+high,−back]	*[mid V]...[V]...[i]	1.1
vocalic	*[−high,−low, +back][ ][+high,+back]	*[o]...[V]...[u]	8.1

Table 18. Shona: unboundedness test

As presented in Table 18, all the placeholder trigram constraints received positive weights on the two different levels. This indicates that these restrictions hold not only across a consonant within a trigram window, but also beyond that window across another vowel, suggesting that these hold at arbitrary distances; these four constraints all passed the unboundedness test.

As the next step, the learner proceeds to investigate if there are blockers for each of these restrictions. Since the evaluation tier for the trigrams learned for Shona is a vowel tier, the middle placeholder of the constraints in Table 17 is replaced by every vowel of the language. Shona has five vowels [a e i o u]; the four placeholder trigrams are now expanded to a set of twenty constraints, as shown in Table 19.

Constraint	Interpretation
*[−high,−low][+low][+high,−back]	*[mid V][a][i]
*[−high,−low][−back,−high][+high,−back]	*[mid V][e][i]
*[−high,−low][−back,+high][+high,−back]	*[mid V][i][i]
*[−high,−low][−low,+back,−high][+high,−back]	*[mid V][o][i]
*[−high,−low][+back,+high][+high,−back]	*[mid V][u][i]
*[+high][+low][−high,−low]	*[high V][a][mid V]
*[+high][−back,−high][−high,−low]	*[high V][e][mid V]
*[+high][−back,+high][−high,−low]	*[high V][i][mid V]
*[+high][−low,+back,−high][−high,−low]	*[high V][o][mid V]
*[+high][+back,+high][−high,−low]	*[high V][u][mid V]
*[+low][+low][−high,−low]	*[a][a][mid V]
*[+low][−back,−high][−high,−low]	*[a][e][mid V]
*[+low][−back,+high][−high,−low]	*[a][i][mid V]
*[+low][−low,+back,−high][−high,−low]	*[a][o][mid V]
*[+low][+back,+high][−high,−low]	*[a][u][mid V]
*[−high,−low, +back][low][+high,+back]	*[o][a][u]
*[−high,−low, +back][−back,−high][+high,+back]	*[o][e][u]
*[−high,−low, +back][−back,+high][+high,+back]	*[o][i][u]
*[−high,−low, +back][−low,+back,−high][+high,+back]	*[o][o][u]
*[−high,−low, +back][+back,+high][+high,+back]	*[o][u][u]

Table 19. Shona: constraints for discovering blockers

These constraints above are reweighted on the vowel tier. The result is summarized in Table 20.

Tier	Constraint	Weight
default	*[any segment]	3.3
Vocalic	*[mid V][a][i]	0
Vocalic	*[mid V][e][i]	1.1
Vocalic	*[mid V][i][i]	9.4
Vocalic	*[mid V][o][i]	0.4
Vocalic	*[mid V][u][i]	0
Vocalic	*[high V][a][mid V]	10.5
Vocalic	*[high V][e][mid V]	10.4
Vocalic	*[high V][i][mid V]	1.8
Vocalic	*[high V][o][mid V]	10.4
Vocalic	*[high V][u][mid V]	10.4
Vocalic	*[a][a][mid V]	1.2
Vocalic	*[a][e][mid V]	9.4
Vocalic	*[a][i][mid V]	9.4
Vocalic	*[a][o][mid V]	9.4
Vocalic	*[a][u][mid V]	9.4
Vocalic	*[o][a][u]	8.8
Vocalic	*[o][e][u]	8.8
Vocalic	*[o][i][u]	8.8
Vocalic	*[o][o][u]	8.8
Vocalic	*[o][u][u]	8.8

Table 20. Shona: discovering blockers

Looking at Table 20, it can be seen that all the five constraints that are derived from either *[high V][ ][mid V], *[a][ ][mid V], or *[o][ ][u] received positive weights. This means that these three vowel restrictions are robust no matter what vowel intervenes, implying that they hold at arbitrary distances and do not decay based on the identity of the intervener. The five constraints that are derived from *[mid V][ ][i] have zero or positive weight. Most importantly, the constraint *[−high,−low][+low][+high,−back] (*[mid V][a][i]) received a zero weight, correctly capturing the role of the low vowel [a] in weakening the restriction of *ei and *oi; the literal interpretation of this result is that the illegal vowel sequences *ei and *oi are no longer prohibited across an intervening [a]. Put differently, the low vowel [a] is a blocker for this specific restriction.

It is noticeable, however, that the constraint *[mid V][ ][i] with an intervening [u] also received the weight of zero, although the high vowel [u] has not been specifically known to be a blocker of this pattern. The current way of discovering blockers is sensitive to how the restriction at evaluation is represented in the grammar, more specifically how granular and/or how violable the restriction is. For instance, the general restriction of Shona *[mid V]...[high V] can be represented differently depending on the learner's tolerance for exception because the two mid vowels do not pattern together with the following high vowel [u]; as summarized previously, whereas both mid vowels cannot precede the high vowel [i] (*[ei], *[oi]), it is only [o] that cannot precede [u] (*[ou]), letting ✓eu occur. If the parameters are set in the way that allows the learner to discover constraints with more violations (soft constraints), the learner will find a violable constraint that describes a behavior of bigger natural class. In fact, in the case study reported by Gouskova and Gallagher (2020), their learner discovered the more general constraint *[−high, −low][ ][+high] ($w = 1.8$), which penalizes *[oi], *[ou], *[ei], as well as the legal ✓[eu] vowel sequence.

Conversely, my learner discovered two constraints that are more specific and have less exceptions instead of a single constraint that is more general and violable, such as *[mid V]...[high V] that the Inductive Projection Learner found. As seen in the baseline grammar for Shona summarized in Table 21, repeated from Table 17, one constraint picks out the two mid vowels as a single natural class that cannot precede the high vowel [i]: *[−high,−low][ ][+high,−back] (*[mid V][ ][i])

($w = 3.8$). And the other picks out only [o] as a natural class that cannot precede [u]: *[−high,−low,+back][ ][+high,+back] (*[o][ ][u]) ($w = 3.9$). Note that these constraints had greater weights than the more violable constraint *[mid V]...[high V] ($w = 1.8$) that the Inductive Projection Learner found.

Constraint	Interpretation	Weight
*[+high][ ][−high,−low]	*[high V][ ][mid V]	5.5
*[+low][ ][−high,−low]	*[a][ ][mid V]	4.5
*[−high,−low][ ][+high,−back]	*[mid V][ ][i]	3.8
*[−high,−low,−back][ ][+high,+back]	*[o][ ][u]	3.9

Table 21. Shona: baseline grammar

The traditional description of Shona height harmony specifies [a] as the blocker segment (Beckman 1997). However, the more specific version of this constraint, *[e]...[i] can actually occur with an intervening [u] because, again, [eu] is an exception to the height harmony and not an illegal vowel sequence. Table 22 shows the result of discovering blockers, repeated from Table 20, with the additional information on the five constraints' observed and expected violations in the learning data.

Tier	Constraint	Observed	Expected	Weight
Vocalic	*[mid V][a][i]	51	3.2	0
Vocalic	*[mid V][e][i]	1	1	1.1
Vocalic	*[mid V][i][i]	0	0	9.6
Vocalic	*[mid V][o][i]	2	2	0.4
Vocalic	*[mid V][u][i]	6	3.1	0

Table 22. Shona: discovering blockers of *[mid V]...[i]

As can be seen in the above table, there were 51 datapoints which included the vowel sequence of [mid V][a][i] in the learning data when only 3.2 instances are expected. There were also 6 examples that violated the constraint *[mid V][u][i] when 3.1 was expected. The 6 instances of [eo][u][i], which are the ones that contributed to the zero weight of *[mid V][u][i], are all e...u...i, which hints to the learner that [u] allows the vowel sequence [e]...[i] to occur.[1]

Hence, the seemingly inaccurate prediction about [u] being a blocker is actually accurate; an intervening [u] in fact licenses the vowel sequence [e]...[i] to occur. The constraint that the learner is finding blockers for, *[mid V]...[i] is a superset constraint of *[e]...[i]. Therefore, the learner identifies both [a] and [u] as blockers of *[mid V]...[i]. My learner is sensitive to granularity of the constraint at evaluation, and the Shona case imposes extra complexity due to the asymmetric behavior of the mid vowels, which are grouped as a single natural class in this particular learning simulation.

Whether or not [u] fits into the category of the traditional blockers, I will assume that [u] will be part of the tier; the smallest natural class to be projected based on the unbounded restriction *[mid V]...[i] already includes [u]. Finally, the tiers that are projected for the Shona learning data are summarized in Table 23. Most importantly, the learner found that the restriction *[−high,−low][ ][+high,−back] can be lifted by an intervening [a] and projected the smallest superset natural class that includes the interacting segments (mid vowels and [i]) and the blocker [a]; this process resulted in the full syllabic tier. The learner did not discover any blockers for the three other restrictions. Hence, only the interacting natural classes for each trigram constraint were projected on three separate tiers.

---

[1]These 6 items include [cheŋudzira] 'to do carelessly', [cheŋurira], [menyukira], [nyeurira], [jeuchidza], [zeŋgurira]. The glosses are partially available on `vashona.com/en/dictionary`.

Constraint	blocker	Tier projection
*[−high,−low][ ][+high,−back]	{a, u}	[+syll]
*[+high][ ][−high,−low]	{ }	[−low]
*[+low][ ][−high,−low]	{ }	[−high]
*[−high,−low,+back][ ][+high,+back]	{ }	[−low, +back]

Table 23. Shona: tier projection

In the final grammar search, the learner discovers constraints on the default tier and on the projected tiers shown in Table 23, looking for both local and non-local restrictions of Shona. The final grammar is presented below.

Tier	Constraint	Interpretation	Weight
−low	*[+high][−high]	*[high V]...[mid V]	5.4
−high	*[+low][−low]	*[a]...[mid V]	5.2
+syll	*[−high,−low][+high,−back]	*[mid V]...[i]	3.6
−low,+back	*[−high][+high]	*[o]...[u]	4.8

Table 24. Shona: final grammar

All the restrictions that were captured as a placeholder trigram constraint in the tier-free search were captured again as their unbounded versions in the final grammar search. Most importantly, the constraint *[−high,−low][+high,−back] was learned on the full syllabic tier instead of on the [−low] tier, successfully penalizing **ei* and **oi* without ruling out ✓*eai* and ✓*oai*.

The learner of Gouskova and Gallagher cannot discover blockers of the pattern, as they project a tier based on a local trigram constraint learned on the default tier, which is not long enough to contain information about blockers. Not being able to find blockers was justified in their paper with the fact that only a small portion of the Shona learning data exhibits such opacity. For example, in their learning data with 4,600 verbal stems, there were only 51 tokens that show the blocking pattern. While it might be not a salient pattern in the entire learning data and capturing it might not improve the fit of the grammar statistically, I believe that the grammar should be still able to capture the role of the blockers if it is psychologically real for Shona speakers; for example, if it is a productive process for them. There is currently no work on testing the psychological reality of such patters among Shona speakers. However, my learner provides a foundational tool to model those patterns, if these turn out to be productive and psychologically real.

4.4 Bi-level distance decay: Arabic OCP

I investigate Arabic place OCP in this section as an example of a language that exhibits distance decay, where a certain phonological restriction gradually peters out with more intervening material (Zymet 2014, Stanton 2016a). As previously mentioned in §3.2, the algorithm proposed in this dissertation and its baseline algorithm by Gouskova and Gallagher (2020) both cannot represent commonly attested patterns of gradual distance decay, where there are significant differences among beyond-transvocalic co-occurrences, such as in Latin *l* dissimilation (Zymet 2014). Instead, my algorithm and the Inductive Projection Learner can capture these patterns as *bi-level decay* at best, with a local trigram constraint and a tier-based bigram constraint which are both positively weighted, as illustrated below.

(34) Representing bi-level decay patterns

	Baseline *b[ ]b $w = 2$	b tier *[−wb][−wb] $w = 2$	$\mathscr{H}$
basis	0	0	0
ba**b**is	-1	-1	-4
basi**b**	0	-1	-2
basisi**b**	0	-1	-2

The tableau in (34) represents a hypothetical language with a *b* co-occurrence restriction. If both the local trigram constraint and the tier-based bigram constraint have positive weights, a stronger version of the restriction (*b[ ]b) holds over one segment and the weaker version of the same restriction holds at a distance that is farther than a segment away, and unboundedly from there without decaying. Looking at the harmony scores in tableau (34), a transvocalic *b* co-occurrence [babis] has the lowest harmony score for violating both constraints. The other two forms, [**b**asi**b**] and [**b**asisi**b**], that have *b* co-occurrences with extra material in between, only violate the tier-based *[−wb][−wb] and therefore end up being more wellformed than the transvocalic *b* co-occurrence (***b****a****b****is*) yet equally illformed and more illformed than the form without any *b* co-occurrence (***b****asis*).

Although my algorithm can only capture the binary distinction between transvocalic and beyond transvocalic co-occurrences, it can be sufficiently explanatory if a language has a word length limitation and thus the decay is mostly only bi-level. This is the case in Arabic. Since most Arabic roots are only triconsonantal and the decay is limited within the root domain, in principle, consonant dependencies in Arabic can be adequately captured as a *bi-level decay* pattern by my learner, assuming sufficient underrepresentation of relevant structures.

Arabic has been extensively studied for its non-concatenative morphological system. Verbal roots mostly consist of a set of two to four consonants, with the most frequent form containing three. Vowels often represent inflectional or derivational information and are inserted between the consonants, conforming to templatic restrictions. For instance, the root /k t b/ 'to write' has its inflected forms *kataba* 'he wrote' and *kutiba* 'it was written', among many others.

Consonants that form a verbal root are under certain co-occurrence restrictions: verbal roots containing homorganic consonants tend to be rare (Greenberg 1950, McCarthy 1986, McCarthy 1994, Frisch et al. 2004, Coetzee and Pater 2008). The traditional description of this restriction discussed by Greenberg and McCarthy is that consonants are divided into natural classes shown in (35), which I call Arabic OCP classes in this section, with co-occurrence restrictions applying within these classes. In their analyses, consonants in any one of these classes can freely co-occur with consonants from any other classes but consonants within any class cannot co-occur. For example, **datam* is highly unlikely since both *d* and *t* belong to the same OCP class (coronal stops). This pattern has been formally captured through OCP-PLACE constraints for each place, such as OCP-LAB, OCP-DOR, and so on (McCarthy 1986, Frisch et al. 2004). Coetzee and Pater (2008) analyze this pattern using a set of constraints that penalize consonant sequences with the same place and/or with the same subsidiary features in CVC configurations; for example, *PP, *PP-SONORANT, *PP-STRICTURE, *PP-VOICE, *PP-EMPHATIC, and *PP-PRENASAL for labial sequences, and so forth for the other places.

(35) Arabic place OCP class

a. Labials = {b f m}

b. Coronal stops = {t d tˁ dˁ}

c. Coronal fricatives = {s z sˁ zˁ ʃ}

d. Coronal sonorants = {n l r}

e. Dorsals = {k g q χ ʁ}

f. Pharyngeals = {χ ʁ ħ ʕ h ʔ} (partially overlaps with dorsals)

The OCP classes in (35) only concern place of articulation, except for the sub-classes of coronals. It has been claimed that other manner features should be taken into consideration for categorizing coronal consonants. McCarthy (1988) claims that the major manner feature [sonorant] should also be considered in classifying Arabic coronals, which separates coronal sonorants from coronal obstruents. Additionally, Padgett (1991) proposes that coronal obstruents are further subdivided into the two categories, coronal stops and coronal fricatives, by the feature [continuant]. The sub-categorization of coronals and the feature specification for each sub-class is shown in Figure 5.

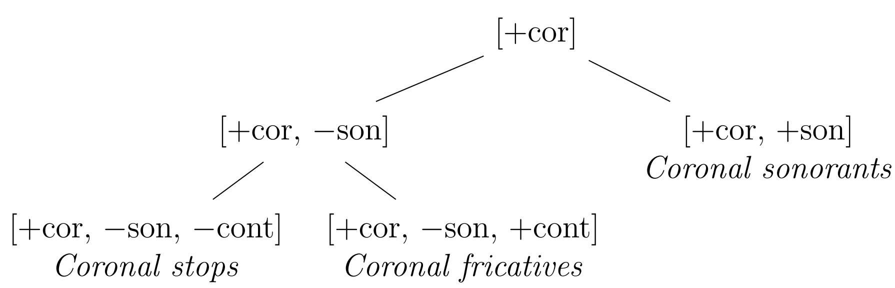

Figure 5. The feature specifications of Arabic coronal OCP classes

Contrary to the traditional description of the pattern by Greenberg and McCarthy, the Arabic place co-occurrence restrictions are gradient and their strengths are correlated to the degree of similarity between the co-occurring consonants (Frisch et al. 2004). For instance, as can be seen in Figure 5, coronal stops and coronal fricatives can be represented as a feature set of [+cor, −son, −cont] and [+cor, −son, +cont], respectively. These two classes share two features [+cor, −son]. Labials can be represented as [+lab] and do not share any feature with coronal stops. Since coronal stops are more similar to coronal fricatives than to labials, the co-occurrence of a coronal stop and a coronal fricative is more avoided than the co-occurrence of a coronal stop and a labial consonant. There is seemingly an exceptional case to this generalization; in fact, many verbs are found with identical C_2 and C_3 consonants. For example, the *r* pair in *farar-* 'flee' seems to violate the strong co-occurrence restriction. However, this results from the autosegmental spreading of the biconsonantal root /f r/ which is motivated by the requirement to fill the template CVCVC, rather than violating the co-occurrence restriction of homorganic consonant. I abstract away from these issues in this study.

Another aspect of the Arabic co-occurrence restriction that has been extensively studied is that it applies more strongly to adjacent consonant pairs (Greenberg 1950, McCarthy 1986, McCarthy 1994, Frisch et al. 2004, and many others). For example, in a triconsonantal root $C_1C_2C_3$, a co-occurrence of homorganic consonants C_1 and C_3 is more likely than a co-occurrence of homorganic consonants C_1 and C_2.

In a nutshell, Arabic place co-occurrence restrictions are gradient and a number of factors influence the strength of each restriction, such as the similarity and the distance between the co-occurring consonants. The restrictions are stronger if the co-occurring consonants are more similar to each other. The restrictions are stronger for adjacent consonant pairs, compared to consonant pairs that have another consonant between them; the Arabic co-occurrence restrictions decay with increasing distance. In the dissertation, I focus only on the distance decay aspect of the Arabic place OCP and abstract away from the similarity factor. Therefore, regardless of the fact that the strength of each co-occurrence restriction is correlated to the similarity between the two co-occurring consonants, I only examine the co-occurrence restrictions that apply within a single OCP class laid out in (35), such as *coronal stop...coronal stop, *coronal fricative...coronal fricative, and *dorsal...dorsal.

The Arabic place co-occurrence restrictions apply only within the root domain and are violated in affixed words (Frisch and Zawaydeh 2001). This is similar to the laryngeal co-occurrence restrictions in Aymara, which is one of the languages that Gouskova and Gallagher (2020) tested their model on. They used a list of roots instead of a word corpus in their Aymara case study (See their §4.2 for more detail). However, I trained the model with a full word list of Egyptian Arabic (Canavan et al. 1997), which contains verbs as well as nouns, many of which have affixes attached to them. For instance, in *ta-driːs* 'NOUN-teaching', the prefix /ta-/ is separate from the root /d r s/ 'to study'. This word is included in the training set even though it has the sequence of homorganic consonants [td] which can serve as counter-evidence to the co-occurrence restriction on coronal stops. Therefore, the evidence for the place co-occurrence restrictions in a word list will be not as salient as in a list that only includes unsuffixed forms.

Most of previous studies that adopt a quantitative approach to the Arabic place co-occurrence restrictions filter out vowels and only examine root consonant sequences (e.g., Frisch et al. 2004). For example, the word *tadriːs* above is considered as its bare root consonants /d r s/. The learner of Gouskova and Gallagher (2020), however, cannot have bare roots as its input data. The learner's success depends heavily on the vowel distributions in the language data, and more cruicially how many CVC trigrams are available to the learner. In order to induce a tier where both [t] and [d] are projected, the learner should be able to first induce a local trigram constraint *t[]d, which is infeasible if the learner was only supplied with bare roots as input.

Since I used a corpus of Egyptian Arabic, some modifications to the data were necessary. For example, any words that contain /p/, /v/, /ʒ/ were excluded from the data since these occur only in loanwords. The training data of Egyptian Arabic included 27,095 words in total. The feature matrix mostly resembled the one presented in Frisch et al. (2004) but some modifications were made in order to facilitate learning. Most importantly, the feature matrix was modified so that the four major places, labial, coronal, dorsal, and pharyngeal, could be picked out as a single feature. For example, pharyngeals are conventionally represented as [+dorsal, +low]. However, pharyngeals needed to be specified as a single feature [+phar] because the learner's heuristics make the learner prefer natural classes that can be described with fewer features. In fact, the learner did not find any relevant pharyngeal constraints when pharyngeals were specified as [+dorsal, +low]. In the Aymara case study reported in Gouskova and Gallagher (2020), similar stipulations were made; plain stops had to be specified as a privative feature [+stop] because the learner failed to include any constraints on plain stops when [−constricted glottis, −spread glottis] was used. Another modification made to the feature matrix is that every consonant was specified as "+" for only a single place feature, except for the two uvular consonants (χ ʁ) and emphatic consonants (tˤ dˤ sˤ zˤ). Uvular fricatives were specified as both [+dorsal, +pharyngeal] since these belong to both dorsal and pharyngeal categories as seen in (35). Also, following Frisch et al. (2004), the emphatic consonants (tˤ dˤ sˤ zˤ) were specified as "+" for both coronal and dorsal.

In order to check how much evidence for place OCP the training data shows, I first looked into the training data if consonant pairs belonging to the same OCP class are underrerpesented. Table 25 shows the O/E ratios of the adjacent consonant pairs (a) and the consonant pairs that have another consonant in-between (b). The way I calculated the O/E ratios is the same as in the Latin case study. In Table 25, the consonant pairs with the matching OCP classes are shaded. Within consonant pairs that do not belong to the same OCP class, the order of consonants was ignored. For example, the value 0.97 for coronal stop and coronal fricative in (a) includes both *dz* and *zd*. At a glance, the corresponding values in (b) are greater than those in (a), indicating that the co-occurrences are more allowed when consonants are not adjacent to each other; the data shows that the co-occurrence restrictions decay with increasing distance. Another thing to note is that the values in Table 25 are much higher compared to the corresponding values in the O/E table presented in Frisch et al. (2004) since my training data included words rather than just roots, and the place co-occurrence restrictions are lifted over a morpheme boundary.

	labial	coronal stop	coronal fric	coronal son	dorsal	phar
labial	0.40					
coronal stop		0.45	0.97	1.00		
coronal fric			0.25	1.06		
coronal son				0.35		
dorsal					0.46	
phar						0.14

a. Adjacent pairs

	labial	coronal stop	coronal fric	coronal son	dorsal	phar
labial	0.82					
coronal stop		0.81	0.94	1.01		
coronal fric			0.61	1.12		
coronal son				1.05		
dorsal					0.77	
phar						0.63

b. Non-adjacent pairs with one intervening consonant

Table 25. Arabic: O/E values of homorganic consonant pairs

Again, it is clear from table (a) above that adjacent homorganic pairs are underrepresented. Although the training data included words like *ta-driːs* ‘teaching’ that can serve as counter-evidence to the place co-occurrence restrictions, there is still evidence that can be provided to the learner that adjacent homorganic pairs are underrepresented.

I first checked if the learner found relevant local trigrams in the baseline grammar. With gain set at 210, placeholder trigrams for the following places are found: labials, coronal fricatives, dorsals, and pharyngeals. The constraints are shown in Table 26.

Constraint	Interpretation	Weight
*[+cont,+phar][ ][+son,+phar]	*[χʁħʕh][ ][χʁħʕhʔ]	2.5
*[−son,+lab][ ][+lab]	*[bf][ ][bfmw]	1.2
*[+cont,+cor,−voice][ ][−son,+cont,+cor]	*[ssˤʃ][ ][szsˤzˤʃ]	1.2
*[+dor,−voice][ ][−son,+dor,+voice]	*[tˤsˤkqχ][ ][dˤzˤg]	1.8

Table 26. Arabic: baseline grammar

As the first part of evaluation, these restrictions are assessed in terms of their unboundedness. Since the interacting natural classes of these trigrams all belong to consonants, the consonantal tier is provided as the evaluation tier. The trigrams shown above are reweighted on this temporarily provided consonantal tier, along with the entire baseline grammar being reweighted on the default tier. As presented in Table 27, these trigrams all gained positive weights on the default tier as well as the temporary consonantal tier. Notably, for each constraint, the weight on the default tier was larger than the weight on the consonantal tier. This result indicates that the restriction holds more strongly over one vowel, and the weaker version of the same restriction is exhibited over an extra consonant; the evaluation result correctly reflects the distance-based decay aspect of Arabic OCP.

Tier	Constraint	Interpretation	Weight
default	*[ ]	*[any segment]	1.7
⋮	⋮	⋮	⋮
default	*[+cont,+phar][ ][+son,+phar]	*[χʁħʕh][ ][χʁħʕhʔ]	2.6
default	*[−son,+lab][ ][+lab]	*[bf][ ][bfmw]	1.2
default	*[+cont,+cor,−voice][ ][−son,+cont,+cor]	*[ssˤʃ][ ][szsˤzˤʃ]	1.2
default	*[+dor,−voice][ ][−son,+dor,+voice]	*[tˤsˤkqχ][ ][dˤzˤg]	1.8
⋮	⋮	⋮	⋮
cons	*[+cont,+phar][ ][+son,+phar]	*[χʁħʕh]...[C]...[χʁħʕhʔ]	0.7
cons	*[−son,+lab][ ][+lab]	*[bf]...[C]...[bfmw]	0.6
cons	*[+cont,+cor,−voice][ ][−son,+cont,+cor]	*[ssˤʃ]...[C]...[szsˤzˤʃ]	0.8
cons	*[+dor,−voice][ ][−son,+dor,+voice]	*[tˤsˤkqχ]...[C]...[dˤzg]	0.9

Table 27. Arabic: unboundedness test

Since these restrictions all passed the unboundedness test, the learner proceeds to look for blockers of each pattern. Each of the constraints shown in (26) are expanded to a set of trigrams, in which the middle placeholder is replaced by the consonants of the language. Egyptian Arabic has nineteen consonants; seventy six trigram constraints are reweighted on the consonantal tier for the sake of discovering blockers. Arabic OCP is not known to have any specific blockers. Conforming to this previous knowledge, all of these constraints received positive weights; the learner found no blockers.

In the evaluation phase, the learner confirms that the four baseline trigrams expand unboundedly and have no blockers. Based on this information, the learner projects four tiers, given in Table 28.

Constraint	Unbounded	blocker	Tier projection
*[−son,+lab][ ][+lab]	✓	{}	[+lab]
*[+cont,+cor,−voice][ ][−son,+cont,+cor]	✓	{}	[−son,+cont,+cor]
*[+dor,−voice][ ][−son,+dor,+voice]	✓	{}	[+dor]
*[+cont,+phar][ ][+son,+phar]	✓	{}	[+son,+phar]

Table 28. Arabic: tier projection

The final grammar for Arabic OCP produced by the Restrictive Tier Learner is summarized by place in Table 29. The learner found multiple bigram constraints on each learned tier, which capture OCP restrictions at arbitrary distances. In addition to these tier-based bigrams, the learner found a general trigram constraint on the default tier, either with a placeholder [ ] or [+syllabic] as a middle gram, which can penalize transvocalic co-occurrences of labials and pharyngeals; thus, for these places, the grammar can successfully represent the OCP patterns as bi-level decay. For other places, namely coronal fricatives and dorsals, no trigram was learned on the default tier. Surprisingly, a trigram that can penalize co-occurrences of coronal sonorants is learned on the default tier.

Tier	Constraint	Interpretation	Weight
+lab	*[+cont][+voice]	*[fw]...[bmw]	1.0
	*[−son,−cont][−wb]	*[b]...[bfmw]	0.5
	*[−son,+cont][−son,−cont]	*[f]...[b]	3.0
default	*[−cont,+lab][+syll][+lab]	*[bm][V][bfmw]	0.8

a. Labial

Tier	Constraint	Interpretation	Weight
−son, +cont, +cor	*[−wb][+acute,−voice]	*[szsˀzˀʃ]...[sʃ]	1.1
	*[−voice][+voice]	*[ssˀʃ]...[zzˀ]	1.8
	*[+acute][−acute,−voice]	*[szʃ]...[sˀ]	2.6

b. Coronal fricatives

Tier	Constraint	Interpretation	Weight
default	*[+son,+cor][−RTR][+son,+cont,+cor]	*[lrn][aiu][lr]	1.5

c. Coronal sonorants

Tier	Constraint	Interpretation	Weight
+dor	*[−son][−wb]	*[tˀdˀsˀzˀkgq]...[tˀdˀsˀzˀkgqχʁɣ]	0.8
	*[−son,−voice][−son,+voice]	*[tˀsˀkq]...[dˀzˀg]	0.9

d. Dorsals

Tier	Constraint	Interpretation	Weight
+son, +phar	*[−wb][+cont]	*[χʁħʕhʔ]...[χʁħʕh]	0.7
	*[+voice][−voice]	*[ʁʕ]...[χħ]	4.8
default	*[+phar][ ][+sonorant,+phar]	*[qχʁħʕhʔ][ ][χʁħʕhʔ]	1.1

e. Pharyngeals

Table 29. Arabic: final grammar

The Restrictive Tier Learner does not do anything differently than the learner of Gouskova and Gallagher when it comes to capturing distance decay patterns in natural languages. More specifically, both algorithms cannot capture gradual distance decay patterns where there are significant difference outside the trigram window. Instead, these learners capture distance-based decay patterns as a bi-level decay pattern, which can be sufficiently adequate if the language lacks long words. Arabic was such a case because most roots are triconsonantal and the OCP is restricted to the root domain, which together make the distance-based decay mostly only bi-level. Similarly to the Quechua case study reported in §4.1, the purpose of presenting the Arabic case study was to show that the inclusion of the evaluation step does not harm learning of distance decay, as opposed to exhibiting how my learner is able to do things differently than the Inductive Projection Learner of Gouskova and Gallagher. With the inclusion of the evaluation step, my learner still finds relevant tier-based bigrams and baseline trigrams, successfully representing the bi-level distance decay of Arabic place OCP.

4.5 Gradual distance decay with blockers: Latin

So far in this chapter, one language was studied to represent each typological pattern: Quechua for unbounded no blocking, Korean for trigram-bound, Shona for blocking, and Arabic for distance decay. As the last case study, Latin is investigated in this section. Latin is a complicated case, as its famous L-dissimilation process exhibits distance decay and blocking, while its phonotactics exhibits trigram-bound place co-occurrence restrictions.

I first briefly present the descriptions of L-dissimilation. In Latin, the realization of the adjectival suffix /-alis/ and the nominal suffix /-al/ largely depends on the presence of *l* in the root (Cser 2007, Cser 2010). With roots that contain no *l*, /-al(is)/ show up faithfully as [-al(is)], as seen in (36a). The /-al(is)/ suffixes surface as their dissimilated forms [-ar(is)] after a root that contains /l/, as seen in (36b−d). This pattern has been traditionally analyzed as dissimilation for the feature [+lateral], where an underlying sequence of *l...l* is mapped to a surface form *l...r*.

(36) Latin L-dissimilation: [-alis] by default but [-aris] for roots with an *l*

a. nav-alis ‘naval’

b. so**l**-aris ‘solar’

c. popu**l**-aris ‘popular’

d. consu**l**-aris ‘consular’

The L-dissimilation is a gradient process whose likelihood depends on the distance between the stem *l* and the suffixal *l* (Zymet 2014 and Stanton 2016a). Examples in (37) below show various distances between the trigger *l* and the target *l*. First, two *l*s barely co-occur when they are in adjacent syllables. Therefore, (37a), for example, almost never surfaces as **mu**l**-a**l**is*, where two *l*s are onsets of adjacent syllables. Second, the dissimilation probability significantly decreases as the two *l*s are farther away from each other, indicating that the *[l]...[l] restriction peters out as more syllables intervene between the two *l*s; *[l]...[l] shows distance-based decay.

(37) L-dissimilation is less likely with more intervening syllables between *l*s

a. mu**l**-aris ‘mules’ (*mu**l**-a**l**is)

b. wu**l**<u>g-a</u>$_\sigma$lis ‘of wheat’

c. di**l**u<u>wi</u>$_\sigma$-<u>a</u>$_\sigma$lis ‘floods’

d. so**l**<u>sti</u>$_\sigma$<u>ti</u>$_\sigma$-<u>a</u>$_\sigma$lis ‘of the summer solstice’

e. **l**ar<u>gi</u>$_\sigma$<u>ti</u>$_\sigma$<u>o</u>$_\sigma$<u>n-a</u>$_\sigma$lis ‘belonging to imperial treasury’

What complicates this pattern even more is that the dissimilation process can be blocked by specific intervening segments. First, a root-final [r] can weaken the restriction (Dressler 1971, Steriade 1987, Bennett 2013, Stanton 2016a and many others). As can be seen in (38), the suffix surfaces faithfully as [-alis] regardless of the presence of [l] in the root, because an [r] intervenes between the trigger and the target [l]. It has been traditionally generalized that an intervening [r] removes the motivation for L-dissimilation.[2]

(38) L-dissimilation blocked by an intervening [r]

a. flor-alis 'floral'

b. later-alis 'lateral'

There are different proposals regarding whether velar and labial consonants can block the L-dissimilation process. In particular, Bennett (2013), McMullin (2016), and Cser (2020) suggest that velar and labial consonants can block the application of L-dissimilation. However, Stanton (2016a) reports a statistical analysis in which the role of non-coronal interveners turns out to be insignificant.

Latin also has a place co-occurrence restriction, in which homorganic consonants tend to avoid each other within syllable boundaries of monosyllabic and disyllabic words (Berkley 2000). These restrictions are most strongly enforced over one segment. For example, Berkley (2000) reports based on a Latin word list extracted from Oxford Text Archive that labial-labial, coronal-coronal, and dorsal-dorsal pairs are underrepresented over one short vowel. She limits the scope of the study to co-occurrence restrictions within a syllable, but she reports that the underrepresentation becomes less significant even within one syllable if there are more intervening segments than a single vowel.

The training data for the Latin case study comes from Gouskova and Gallagher (2020) and includes 84,046 words of Latin. I use the entire database as a learning data although the L-dissimilation is only observed in a sublexicon of Latin, namely

[2] Stanton (2016a) provides an analysis from a different perspective, arguing that the failure of L-dissimilation on roots with an intervening [r] is evidence for a [r] co-occurrence restriction. The reason why the suffixal [l] faithfully surfaces is not because the intervening [r] blocks the application of L-dissimilation. Rather, it faithfully surfaces because it would otherwise violate the [r] co-occurrence restriction, as in *[flo**r**-a**r**is] and *[late**r**-a**r**is], which is worse than violating **l...l* within a word.

the words that end in [-alis], [-aris], [-al], or [-ar], because I also wanted the learner to discover constraints that capture the place OCP restrictions. Gouskova and Gallagher (2020) point out that supplying the relevant sublexicon directly to the learner facilitates the learning process because the learner is provided with the more concentrated evidence for certain patterns. For example, in their learning simulations for Shona in which the vowel co-occurrence restrictions hold within the verbal stems, the learner did not find any relevant constraints when trained with the entire word corpus that includes nouns and morphologically complex forms of verbs. The learner was able to find relevant placeholder trigrams and properly project tiers when trained only with the verbal stems. Thus, it is expected that discovering the desired constraints from the entire lexicon of Latin will be more challenging.

The Latin feature matrix also comes from Gouskova and Gallagher (2020). It was included along with the Latin corpus in their learner file. I shortened all the long vowels and geminates in the corpus to ease the learning process and thus also removed the [long] feature from their feature chart. Everything else remained as it was.

The present paper focuses on the **l...l* that is exhibited in the form of L-dissimilation and place OCP that is exhibited as trigram-bound phonotactic restrictions, such as *[+labial][ ][+labial], *[+coronal[[ ][+coronal], and *[+dorsal][ ][+dorsal]. I first check if the above pairs are underrepresented at different distances in the learning data. In Zymet (2014) and Stanton (2016a), the unit of measuring the distance is the number of syllables. In the current study, the number of intervening consonants is used as a distance measure to avoid extra complexity that comes from syllabification.

	l
l	0.13

a. **C**V.**C**
transvocalic

	l
l	0.54

b. **C**(V)C(V)**C**
1 consonant away

	l
l	0.72

c. **C**(V)C(V)C(V)**C**
2 consonants away

Table 30. Latin: O/E values of *l...l* at different distances

Table 30 shows the O/E ratios of *l* co-occurrences with no intervening consonant (a), one intervening consonant (b), and two (c). The table in (a) shows the O/E ratio of *l* co-occurrences when there is a single vowel between them. As can be clearly seen in the table, *l* pairs are highly underrepresented when they are only a vowel apart. The table in (b) shows the O/E ratios of *l* co-occurrences when there is one intervening consonant between them; the O/E values indicate that a pair of two *l*s are still underrepresented (0.54) when one consonant intervenes. The table in (c) shows the O/E ratio of *l* co-occurrence when there are two intervening consonants between them. Similarly, the *l* co-occurrences are still somewhat underrepresented (0.72). Based on the observation made from Table 30, we can conclude that the *l* co-occurrence restriction is gradually loosened as the number of intervening consonants monotonically increases, which is compatible with the previous studies (Zymet 2014, Stanton 2016a), which used the number of intervening syllables as a distance measure.[3]

Subsequently, I computed the O/E values of labial-labial, coronal-coronal, and dorsal-dorsal pairs in the learning data at the following distances: transvocalic, one consonant away, and two consonants away. Berkley (2000) discusses how the distinction between obstruents and sonorants for coronals is unnecessary in analyzing Latin

[3]In fact, Zymet (2014) shows that the syllable count and the segment count are highly correlated in Latin ($r = 0.95$); the consonant count measure does not misrepresent the previous generalizations.

place OCP. For simplicity, I aggregate both obstruents and sonorants within a place; for example, labials (p, b, f, m, w), coronals (t, d, s, n, l, r), and dorsals (k, g).

	lab	cor	dor
lab	0.59		
cor		0.91	
dor			0.19

a. **C**V.**C**
transvocalic

	lab	cor	dor
lab	0.92		
cor		1	
dor			1

b. **C**(V)C(V)**C**
1 consonant away

	lab	cor	dor
lab	0.94		
cor		1	
dor			1

c. **C**(V)C(V)C(V)**C**
2 consonants away

Table 31. Latin: O/E values of homorganic pairs at different distances

As shown in Table 31 (a), the learning data shows underrepresentation of labial-labial and dorsal-dorsal pairs within a trigram window. When there is one or more intervening consonant, as shown in (b−c), those pairs are not underrepresented anymore. Coronal co-occurrences are not specifically underrepresented at any distance, which is a different observation than Berkley (2000). I will only focus on the underrepresentation of labial and dorsal pairs moving forward.

The underrepresentation of transvocalic labial and dorsal co-occurrences can be captured via local trigram constraints such as *[+lab][ ][+lab] and *[+dor][ ][+dor] on the default tier; a tier projection from these constraints is not necessary. Conversely, capturing the distribution of [l]s can benefit from having tier-based representations. The **l...l* restriction peters out as the number of intervening consonants monotonically increases; and a successful grammar of Latin should capture this distributional restrictions of [l]s in the following way. First, the nearly inviolable restriction *l[ ]l (such as in *so**l**-a**l**is) should be picked up as the local trigram constraint *l[ ]l during the baseline grammar search. Second, the trigram constraint *l[ ]l should motivate a tier with only [l]s and other relevant segments, such as blockers. On this tier, there should be a constraint *[−wb][−wb], which penalizes a co-occurrence of *l*s, no matter what the distance between these two *l*s is. If the final grammar also finds a local trigram that can penalize transvocalic [l] co-occurrences, on top of the tier-based bigram, the final grammar can represent the [l] co-occurrence restriction as a bi-level distance decay pattern.

The scenario described above is the ideal way of representing a gradual distance decay for my algorithm. However, this grammar will not be able to capture every shade of the **l...l* decay due to the nature of tier representations. For instance, the [l] tier, whose existence is crucial for penalizing the beyond-transvocalic [l] co-occurrences, is blind to the distance between the two co-occurring [l]s. The default (baseline) and the tier-based representations for hypothetical [l] co-occurrences are displayed in (39). As can be seen, no matter how many consonants occur between the two *l*s, these are all represented as a local *ll* sequence on the *l* tier. Therefore, even though /**l**aposa**l**is/ is more wellformed than /pa**l**osa**l**is/ for having an extra intervening consonant, this increased wellformedness would not be captured through the current learning model. Put differently, every degree of distance decay cannot be captured in a grammar where local restrictions are captured only up to a trisegmental distance and restrictions at further distances are captured through tier-based constraints.

(39) Default and tier-based representations for noncewords with *l* co-occurrences

l tier	l l	l l	l l
Baseline (all segs)	pasolalis	palosalis	laposalis

The ideal, though not fully fine-grained, final grammar for Latin L-dissimilation is illustrated in (40). The first three nonce word forms in (40) represent words with different distances between two *l*s: *paso**l**a**l**is* for the cases where *l*s are transvocalic, and *pa**l**osa**l**is* and ***l**aposa**l**is* for beyond transvocalic. In each form, the boldfaced segments are the ones that are visible on the *l* tier. The first form *paso**l**a**l**is* is penalized by both the local trigram constraint *l[ ]l and the tier-based constraint

*[−wb][−wb]. The other two forms *pa**l**osa**l**is* and ***l**aposa**l**is*, on the other hand, are penalized by the tier-based constraint *[−wb][−wb] and not by *l[ ]l, since there is more than one segment between the two *l*s. As mentioned above, the increased wellformedness of ***l**aposa**l**is*, compared to *pa**l**osa**l**is*, will not be captured since the two forms have the exact same violation profile. Therefore, the grammar will be able to make a binary distinction between transvocalic and beyond-transvocalic *l* co-occurrences but will not be able to capture the difference between the two beyond-transvocalic ones. Regarding the weighting condition of these two constraints, the following is anticipated. The weight of the tier-based constraint *[−wb][−wb] should be adjusted to reflect the badness of words that have two *l*s, no matter what the distance between them is. The weight of *l[ ]l, on the other hand, should reflect how much *worse* the transvocalic *l* co-occurrence is, compared to *l* co-occurrences at any distance.

(40) The ideal grammar for Latin

l tier		*ll		
baseline	*l[ ]l		*[+lab][ ][+lab]	*[+dor][ ][+dor]
paso**l**a**l**is	-1	-1		
pa**l**osa**l**is		-1		
laposa**l**is		-1		
bibasis			-1	
probakis				
kokimaris				-1
kalkamis				

Regarding the transvocalic place co-occurrence restrictions, the following is anticipated. The last four nonce word forms represent words with homorganic pairs at transvocalic and beyond-transvocalic distances. If the trigram constraints on the default tier, *[+lab][ ][+lab] and *[+dor][ ][+dor], are learned in the baseline grammar search, the evaluation should be able to rule out the tier projection based on them. In the final grammar search, these two trigrams should be discovered again in order to capture the trigram-bound co-occurrence restrictions on labial and dorsal pairs.

I first report the learning simulation results produced by the Inductive Projection Learner. With the gain value of 100 and gamma of 0, the Inductive Projection Learner found numerous placeholder trigrams in the baseline grammar search, as summarized in Table 32. The use of moderate parameters were necessary to discover the trigram-bound restriction of dorsal OCP (e.g., it could not be higher) but it also led to discovering many other constraints that later cause a problem.

Constraint	Interpretation	Weight
*[+dorsal][ ][+dorsal]	*[kg][ ][kg]	0.9
*[+lateral][ ][+lateral]	*[l][ ][l]	1.9
*[−cont,+lab][ ][−son,+cont]	*[pbm][ ][fsh]	0.9
*[−son,−cont,+lab][ ][+son,−cont]	*[pb][ ][mn]	0.6
*[−cont,+voice,+cor][ ][−cont,+voice]	*[dn][ ][bmgdn]	0.6
*[−son,+cont][ ][−cont,+cor]	*[fsh][ ][tdn]	0.3
*[−son,+voice][ ][−son,+voice,+cor]	*[bdg][ ][d]	0.8

Table 32. Latin: baseline grammar by the Inductive Projection Learner

The Inductive Projection Learner successfully discovered the desired trigram constraints: *[+dorsal][ ][+dorsal] and *[+lateral][ ][+lateral]. The learner also found several other placeholder trigram constraints that seem trivial; they cannot be generalized as a naturalistic restriction. Regardless, based on these trigram constraints, the Inductive Projection Learner automatically projects tiers shown in Table 33, on which constraints are discovered.

Constraint	Tier projection
*[+dorsal][ ][+dorsal]	[+dorsal]
*[+lateral][ ][+lateral]	[+lateral]
*[−cont,+lab][ ][−son,+cont]	[−cont]
*[−son,−cont,+lab][ ][+son,−cont], *[−son,+cont][ ][−cont,+cor]	[+cons]
*[−cont,+voice,+cor][ ][−cont,+voice]	[−cont,+voice]
*[−son,+voice][ ][−son,+voice,+cor]	[−son,+voice]

Table 33. Latin: tier projection by the Inductive Projection Learner

The final grammar produced by the Inductive Projection Learner is summarized in Table 34. In fact, in order for the program to learn the final grammar, the maximum length for tier-based constraints had to be adjusted down from its default value three to two, because the learner ran out of memory and crashed without this modification. Limiting the upper bound of constraint length to two substantially reduced the search space, allowing the program to run successfully.

Tier	Constraint	Interpretation	Weight
+cons	*[+lateral][+lateral]	*[l]...[l]	1.0
+cons	*[+dor][+dor,+voice]	*[kg]...[g]	0.9
+cons	*[+dor][+dor,−voice]	*[kg]...[k]	0.5
+cons	*[+voice,+lab][+son,+lab]	*[pbm]...[pbfw]	1.1
default	*[+dor][−high,−back][+dor]	*[kg][e][kg]	3.0
default	*[−cont,+lab][−high,−back][−nasal,+lab]	*[pbm][e][pbfw]	1.4

Table 34. Latin: final grammar by the Inductive Projection Learner

The constraint that can penalize [l] pairs at any distance, such as *[−wb][−wb] on the lateral tier, was initially learned but the eventual weight of it was adjusted to zero. Instead, *[+lateral][+lateral] was learned on the [+consonantal] tier. As this restriction is captured on the consonantal tier, it can only penalize [l] co-occurrences across vowels or/and glides, and not across other consonants. Overall, the final grammar fails to represent the [l] distributions as a bi-level decay, as there is no constraint that can penalize [l] pairs that occur at arbitrary distances, across syllables with an onset consonant. The transvocalic place OCP was successfully captured for labials and dorsals, as multiple constraints. As summarized in Table 34, these phonotactic restrictions were learned on the consonantal tier as a bigram as well as on the default tier as a trigram.

To summarize, while the learner of Gouskova and Gallagher successfully captured the trigram-bound phonotactic restrictions, the unavoidable gain choice to enable it led the learn to projecting too many tiers. These unnecessary tiers distracted the learner in the final grammar search, and the learner failed to capture the gradual decay of *[l]...[l] as a bi-level decay pattern.

Now I report the learning simulation results produced by the Restrictive Tier Learner. I will show how evaluation can prevent unnecessary tiers from being projected. With the same set of parameters (100 gain and 0 gamma), the learner discovered the following placeholder trigrams in the baseline grammar search, summarized in Table 35.

Constraint	Interpretation	Weight
*[+dorsal][ ][+dorsal]	*[kg][ ][kg]	1.3
*[−son,−cont,+lab][ ][+son,−cont]	*[pb][ ][mn]	0.8
*[−cont,+lab][ ][−son,+cont]	*[pbm][ ][fsh]	0.7
*[+lateral][ ][+lateral]	*[l][ ][l]	1.7
*[−son,+cont][ ][−syll,−cons]	*[fsh][ ][wj]	0.8
*[−cont][ ][−son,+voice]	*[pbmkgtdn][ ][bdg]	0.4

Table 35. Latin: baseline grammar

Instead of directly projecting a tier from each trigram constraint shown above, the learner first determines if the restriction captured as these trigrams actually generalize unboundedly. The above restrictions are all consonantal interaction; these constraints are reweighted on the evaluation C tier along with the entire grammar on the default tier. The result of the unboundedness test is shown in Table 36.

Tier	Constraint	Interpretation	Weight
default	*[ ]	*[any segment]	0
⋮	⋮	⋮	⋮
default	*[+dorsal][ ][+dorsal]	*[kg][ ][kg]	1.3
default	*[−son,−cont,+lab][ ][+son,−cont]	*[pb][ ][mn]	0.8
default	*[−cont,+lab][ ][−son,+cont]	*[pbm][ ][fsh]	0.7
default	*[+lateral][ ][+lateral]	*[l][ ][l]	1.7
default	*[−son,+cont][ ][−cons]	*[fsh][ ][wj]	0.8
default	*[−cont][ ][−son,+voice]	*[pbmkgtdn][ ][bdg]	0.3
⋮	⋮	⋮	⋮
cons	*[+dorsal][ ][+dorsal]	*[kg]...[C]...[kg]	0
cons	*[−son,−cont,+lab][ ][+son,−cont]	*[pb]...[C]...[mn]	0
cons	*[−cont,+lab][ ][−son,+cont]	*[pbm]...[C]...[fsh]	0
cons	*[+lateral][ ][+lateral]	*[l]...[C]...[l]	0.4
cons	*[−son,+cont][ ][−cons]	*[fsh]...[C]...[wj]	0
cons	*[−cont][ ][−son,+voice]	*[pbmkgtdn]...[C]...[bdg]	0

Table 36. Latin: unboundedness test

As can be seen in the above table, the placeholder trigrams discovered in the baseline grammar search are mostly only trigram bound; except for *[+lateral][ ][+lateral] that is already known to us to be unbounded, all the constraints are weighted zero on the consonantal tier. This is the part where my learner performs differently than the Inductive Projection Learner. The Restrictive Tier Learner does not project a tier if the restriction captured in a placeholder trigram is only valid within a trigram window and does not hold outside of it.

As the second part of evaluation, the Restrictive Tier Learner proceeds to discover blockers of the restriction *[l]...[l], if there are any. The trigram constraint *[+lateral][ ][+lateral] is expanded to a set of trigrams in which the middle placeholder is replaced by each one of all the consonants in Latin. Latin has 14 consonants; the 14 constraints are reweighted on the consonantal tier, as shown in Table 37.

Tier	Constraint	Interpretation	Weight
cons	*[+lateral][−cont,−voice,+lab][+lateral]	*[l][p][l]	1.2
cons	*[+lateral][−son,+voice,+lab][+lateral]	*[l][b][l]	0
cons	*[+lateral][+cons,+son,+lab][+lateral]	*[l][m][l]	1.2
cons	*[+lateral][+cons,+cont,+lab][+lateral]	*[l][f][l]	1.6
cons	*[+lateral][−cons,+lab][+lateral]	*[l][w][l]	0.5
cons	*[+lateral][−voice,+dor][+lateral]	*[l][k][l]	0
cons	*[+lateral][+voice,+dor][+lateral]	*[l][g][l]	1.9
cons	*[+lateral][−cont,−voice,+cor][+lateral]	*[l][t][l]	0.1
cons	*[+lateral][−son,+voice,+cor][+lateral]	*[l][d][l]	2.7
cons	*[+lateral][−son,+cont,+cor][+lateral]	*[l][s][l]	1.2
cons	*[+lateral][+son,−cont,+cor][+lateral]	*[l][n][l]	1.6
cons	*[+lateral][−lateral][+lateral]	*[l][r][l]	1.0
cons	*[+lateral][+lateral][+lateral]	*[l][l][l]	12.2
cons	*[+lateral][−cons,+cor][+lateral]	*[l][j][l]	12.3

Table 37. Latin: discovering blockers

As summarized at the beginning of this Latin section, the segments that have been reported to be blockers of the L-dissimilation pattern are most frequently claimed [r]

(Dressler 1971, Steriade 1987, Bennett 2013, Stanton 2016a and many others) as well as non-coronal consonants (Cser 2010, Cser 2020, Bennett 2013, McMullin 2016). The learning results shown above do not provide clear-cut evidence for any of these previous observations. Most importantly, the constraint with [r] in the middle, was not weighted zero, which indicates that the sequence of l...r...l is not necessarily overrepresented in the learning data. The only constraints that received zero weights are the ones with [b] and [k] in the middle. Those are non-coronal consonants, which is compatible with the previous generalization that labial and velar consonants also block dissimilation. However, not all the non-coronal consonants received zero either, which further complicates the interpretation of this result.

The current version of the Restrictive Tier Learner cannot project a tier that does not form a natural class. The smallest natural class that includes the interacting segment [l] and the discovered blockers [b, k] is [+cons, −nasal], which includes all the consonants except for glides and nasals. While this tier is not restricted only to [r] and the non-coronal blockers, I proceed with the projection of this tier for now as it includes all the segments that need to be visible on the single tier.

In the final grammar search, the Restrictive Tier Learner finds the following constraints, summarized below.

Tier	Constraint	Interpretation	Weight
+cons,−nasal	*[+lateral][+lateral]	*[l]...[l]	1.4
default	*[+dor][ ][+dor]	*[kg][ ][kg]	0.8
default	*[+lab][ ][−cont,+lab]	*[pbmfw][ ][pbm]	0.9
default	*[+cons,+son,+cor][−low][+son,+cor]	*[nrl][ieou][nrlj]	0.5

Table 38. Laitn: final grammar

The bigram constraint *[+lateral][+lateral] was learned on the [+cons,−nasal] tier, which can penalize some [l] pairs that occur at various distances; it can penalize transvocalic [l] pairs as in lVl, and it can also penalize co-occurrences over one or more syllables, as long as it is a syllable with a glide or a nasal consonant. Unfortunately, the constraint cannot capture [l] pairs that occur across syllables with non-blocking consonants, such as coronal obstruents. Additionally, the learner found the constraint *[+cons,+son,+cor][−low][+son,+cor] on the default tier, which can extra penalize transvocalic [l] co-occurrences, capturing the distance decay aspect of the restriction *[l]...[l] as a bi-level distance decay. As can be seen in Table 38, the transvocalic place co-occurrence restrictions were captured for labials and dorsals, as placeholder trigram constraints on the default tier.

To summarize, the Restrictive Tier Learner shows better performance on the Latin data, by having an extra evaluation step in which a tier projection is prevented if the placeholder trigram is only trigram-bound or trivial. The trigram-bound place co-occurrence restrictions are initially learned in the baseline grammar search but the evaluation step rules them out from a tier projection. The learner does not get distracted by a handful of unnecessary tiers in the final grammar search, which helps it to successfully discover these restrictions as placeholder trigrams on the default tier again. Only the necessary tiers are projected in the final grammar search, on which unbounded *[l]...[l] can be captured. While discovering blockers for *[l]...[l] is only partially successful, it still discovers a tier-based bigram and a local trigram, capturing the distance decay aspect of the restriction.

4.6 Summary

In this chapter, I presented five case studies to test the abilities of the Restrictive Tier Learner on capturing various long-distance dependencies that are attested in natural languages. The findings are as follows.

The main difference that my learner has is the ability to distinguish unbounded patterns from trigram-bound ones. In my learner, restrictions that are captured as a local trigram are not carried over for a tier projection unless they pass the unboundedness test in the evaluation phase. For the Korean case study in §4.2, the local trigram *[+dor][ ][+dor] did not pass the unboundedness test because the restriction is merely trigram-bound. In the Latin case study reported in §4.5, several trigrams were ruled out in the evaluation because they were either trigram-bound or junk constraints, meaning that they just happened to be true generalizations (Albright and Hayes 2006). For a comparison, the learning result of the Inductive Projection Learner was presented in parallel, which resulted in a projection of the unnecessary tiers and a failure in discovering a trigram constraint that is actually necessary. The projection of the unnecessary tiers distracted the learner and made the learner discover several tier-based constraints that are trivial while not discovering the necessary local trigram in the Korean case study. In the Latin simulations, automatic tier projection resulted in the projection of too many tiers, causing the learner to run out of memory and crash.

There are cases in which the Restrictive Tier Learner does not perform any differently than the previous learning algorithm of Gouskova and Gallagher. The unbounded no decay case and the distance decay case belong here. For these cases, only the learning results produced by the Restrictive Tier learner were presented, to show how the addition of the evaluation phase does not harm the learning process. As an unbounded no decay example, the phonotactics of Quechua laryngeals were examined in §4.1. Cases like this can be already smoothly learned by the learner of Gouskova and Gallagher. The results of Restrictive Tier Learner showed that the addition of evaluation still allows the necessary tiers to be projected. Both the Inductive Projection Learner and my learner cannot represent the attested patterns of gradual distance decay; these patterns can be represented as a bi-level decay at most.

Although these learners cannot capture every shade of gradual distance decay, it can be sufficient if the language lacks long words. Arabic is the case of this as its roots are mostly triconsonantal and the restriction is bounded to the root domain. The learning result given in §4.4 showed that my learner can successfully represent bi-level distance decay for some places by discovering a tier-based bigram and a local trigram.

The Restrictive Tier Learner was not always successful in discovering blockers. In the case study on Arabic place OCP, which is not specifically known to have blockers, the learner did not pick out any blocker, as desired. In the Shona case, the blockers that were picked out by the learner did not perfectly match with the traditional description of the pattern; but the pattern was actually successfully captured. In the Latin case study presented in §4.5, the learner picked out not enough blockers; the learner did not find [r] as a blocker although it is most frequently claimed blocker of the pattern, while discovering some, but not all, of the non-coronal consonants. The current learner projects the smallest superset natural class that includes all the discovered blockers. Thus, not picking out the exact set of blockers did not change the natural class of projection in both case studies. In the future, though, if non-natural class tiers are proven to be necessary by experimental work, the learner should be able to identify the exact set of blockers and only those.

To conclude, the current version of the learner does not deteriorate what the previous learning algorithms do while showing improved performance on some patterns. Although discovering blockers requires further research for more stable learning, the partial success shown in Shona and Latin is a good first step.

CHAPTER 5
DISCUSSION

My learner makes certain predictions about typology and phonological tiers. My learner also has limitations that arise from assumptions that my learner is built upon. In this chapter, I provide in-depth discussions on these predictions and assumptions. In §5.1, I discuss inherent learning challenges that my learner faces in dealing with trigram-bound patterns, in connection with the typological rarity of such patterns. In §5.2, I compare general tiers and specific tiers, and the learnability of those in different approaches. My learner cannot represent interactions between consonants and vowels; the discussion around this issue is given in §5.3. My algorithm relies on the universal preference toward simple syllable structures, capturing bounded patterns only with local trigram constraints; the discussion about this can be found in §5.4. Lastly, my learner can only deal with tiers that form a natural class. I discuss the possibility of positing non-natural tiers in §5.5. I summarize the chapter in §5.6.

5.1 Learnability of trigram-bound patterns

Throughout the dissertation, I noted how trigram-bound patterns bear inherent learning challenges. Due to the nature of being bounded, these patterns create systemic exceptions in the learning data and obscure the pattern as an overall phonotactic in a language. Thus, learning of these languages requires special treatment so that such patterns can still be discovered by the learner.

First, hyperparameters can be adjusted to make the learner more lenient with exceptions. For example, in the Korean case study reported in §4.2, gain values were adjusted down to discover a trigram-bound restriction of the language. The learner was able to find the relevant trigram constraint *[+dor][][+dor] with the gain at 15 and the gamma at 0. Second, if it is the case that a trigram-bound pattern holds only within a specific sublexicon of a language, supplying the learner directly with that portion of the data can aid learning, because, as Gouskova and Gallagher (2020) point out, the learner is provided with more concentrated evidence for certain patterns. But even so, if the language contains a lot of long words, it is unavoidable that the learner will encounter counterevidence. Lastly, if the trigram bound pattern can be referenced to a specific location in a word, such as a morpheme boundary or a word boundary, the restriction might become more discoverable. Although not specific to bounded cases, Gallagher et al. (2019) already showed that morphologically parsed learning data can be beneficial. In their paper, they test the Inductive Projection Learner's ability to capture Aymara laryngeal phonotactics, which are very similar to the ones exhibited in Quechua but crucially different in the sense that the restrictions *plain stop...ejective and *plain stop...aspirate only hold within a root; these can be violated over a morpheme boundary. The learner successfully discovered constraints such as *[plain stop][−morpheme boundary][ejective/aspirate] from the morphologically parsed learning data such as [qaq+t^{h}api+ɲa] 'to finish scratching', correctly generalizing the lexical statistics observed within the root domain. Thus, the hope here is that bounded patterns might be made more discoverable with the help of extra morphological information. McCollum's work discusses how some attested cases of non-iterative harmony can be initiated anywhere in a word and are not dependent on the word edge; thus, the morphological information cannot be of help in these cases. However, Lamba exhibits a trigram-bound nasal harmony in a form of suffixal alternation; among many, the perfective suffix /-ile/ surfaces as [-ine] if a nasal consonant precedes across a single vowel, and otherwise surfaces faithfully. If the morphological information is incorporated into either the feature matrix or into the learning data, these patterns might be learned more smoothly.

To conclude, learning bounded patterns requires *some* mechanism to be more complex. It can either lie in the representations or in the grammar (e.g, multiple-grammar approach as in Allen and Becker 2015). Relatedly, the fact that trigram-bound patterns are a challenge for a learner is probably the reason why bounded patterns are rare across and within languages. Bounded patterns are typologically rare, or limited to a specific sublexicon even if a language exhibits such a pattern. For example, Korean place OCP is stronger in monosyllabic words and Latin place OCP is limited to monomorphemic words that are monosyllaic or disyllabic. The nasal harmony of Lamba is only specific to certain suffixes. Based on these observations, it can be safely assumed that learners need to rely on some cue to refer to, whether a sublexicon-specific grammar or a representational marker, in order to learn a bounded pattern in most cases.

There is recently a growing body of research that points out the connection between typology and learnability, and more specifically how learnability can shape typology (Stanton 2016b). McCollum and Kavitskaya (2022) discuss the difficulty of modeling non-iterative harmony within constraint-based theories because the harmony domain is not always aligned to a specific morphoprosodic edge.[1] Experimental work by Finley (2011), Finley (2012), and Jardine and McMullin (2017) shows that unbounded patterns are easier to learn. The learning challenges that I showed in Korean and Latin case studies can also add to empirical evidence that bounded patterns are hard to learn, making the claim stronger about how learanbility influences typology.

[1]For instance, if it was the case that the harmony trigger of Crimean Tatar was always the word initial vowel, the pattern could have been represented as *[+wb][+round][−round], on a tier with vowels and the word boundary.

5.2 General tier vs. Specific tier

The description of phonological long-distance dependencies can either be achieved on the tier level or on the constraint level; as schematized in Figure 6, there is a specificity trade-off between tiers and constraints. For instance, the height harmony in Shona *[mid V]...[high V], can be represented on the general syllabic tier as *[−high,−low][+high]. On the other side of the spectrum, there can also be a scenario where the a very similar constraint is represented as *[−high][+high] on the [−low] tier.

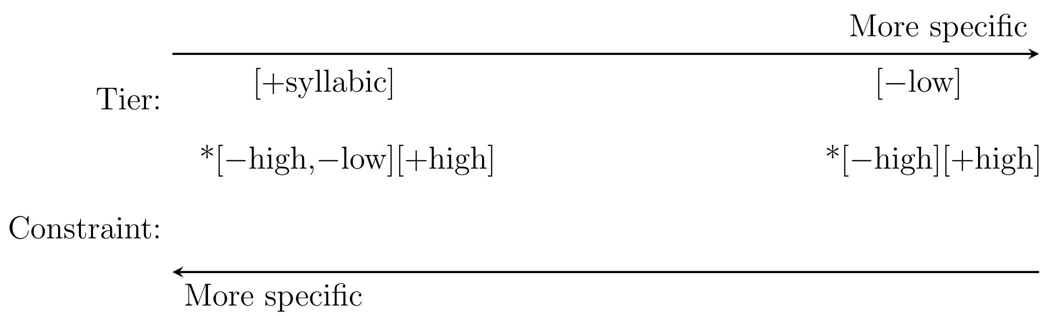

Figure 6. Specificity spectrum of *[mid V]...[high V]

The existing learners of Hayes and Wilson (2008) and Gouskova and Gallagher (2020) employ different specificity levels of the tier on which Shona height harmony is represented. In Hayes and Wilson (2008), tiers need to be defined and supplied to the learner by the analyst. Thus, the learner was directly provided with the general [+syllabic] tier in the Shona simulation, which is conventional for capturing vowel harmony. The learner of Gouskova and Gallagher (2020) can induce tiers automatically, based on evidence that is visible in local trigrams. In the Shona simulation reported in Gouskova and Gallagher (2020), their learner projected multiple tiers ([−high], [−low], and [−high, −low]), instead of a single [+syllabic] tier. The above restriction was captured as *[−high][+high] on the automatically induced [−low] tier. The two approaches above have their own advantages. The Inductive Projection Learner has an advantage of requiring less input from the analyst and inducing necessary tiers on its own. By contrast, as already pointed out by Gouskova and Gallagher (2020), the [+syllabic] tier, which is what linguists are more likely to propose to analyze a vowel harmony pattern, can provide a more accurate description of the language in dealing with blockers. For example, as was shown in §4.3, the otherwise illicit sequences, *e...i and *o...i, become grammartical if a blocker vowel *a* intervenes. This blocking pattern can be captured only on a tier where every vowel (low, mid, and high), is visible.

In the Shona simulation that Gouskova and Gallagher (2020) report, the reason why the learner did not induce a [+syllabic] tier is because there was no placeholder trigram that could lead to a projection of a general [+syllabic] tier. Their learner projects the smallest superset natural class that includes either side of a placeholder trigram; the baseline grammar included *[−high,−low][][+high], which led to the projection of a tier that excludes the low vowel [a]. Since my learner is equipped with an extra evaluation step where blockers are discovered, my learner was able to automatically induce the [+syllabic] tier where not only the interacting non-low vowels but also the opaque low vowel is visible. Table 39 is a comparison between the previous learners and mine; Hayes and Wilson (2008) has an ability to represent blockers but requires an input from the analyst. Gouskova and Gallagher (2020) has an advantage of being able to induce a tier while not being able to represent opaque patterns; a placeholder trigram, which their learner projects a tier from, is too short to contain information about blockers. My learner is able to do both.

	H&W (2008)	G&G (2020)	This dissertation
Induce tiers		✓	✓
Represent blockers	✓		✓

Table 39. Comparing phonotactic learners

In theory, long-distance consonantal dependencies can also be represented along the specificity spectrum. For instance, unbounded place OCP in Arabic, such as *[+cont, +cor]...[+cont, +cor], can be expressed on the specific [+cont, +cor] tier as a general constraint *[−wb][−wb], shown on the right side of the spectrum in Figure 7. Alternatively, the specificity can be biased toward the constraint level; the restriction can be expressed on the more general tier [+consonantal] as *[+cont, +cor][+cont, +cor]. In the middle of the specificity spectrum, there can also be a scenario where the restriction is discovered on the more general [+cor] tier as *[+cont][+cont], which adds a manner feature on top of the place information that is expressed on the tier level.

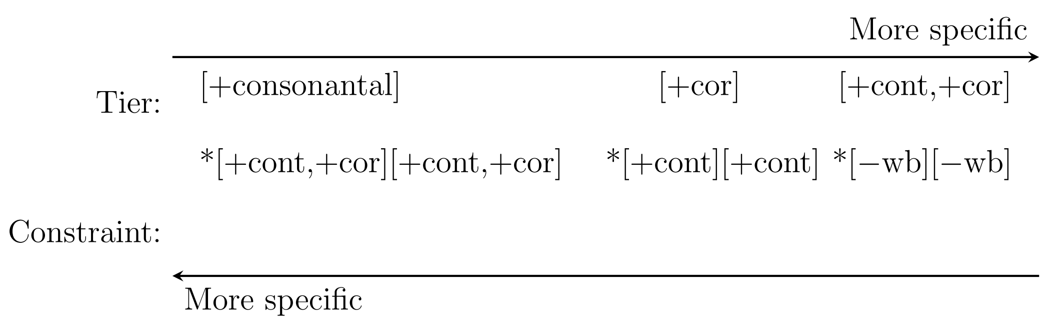

Figure 7. Specificity spectrum of *[+cont, +cor]...[+cont, +cor]

In fact, the [+coronal] tier cannot be induced from the current learning data of Arabic OCP. Arabic has three subcategories of coronal, which all form independent OCP classes from one another. If the final grammar wants to have a [+coronal] tier, the baseline grammar should discover a constraint such as *[+cor+son][ ][+cor−son] such that the smallest natural class that includes [+cor+son] and [+cor−son] is [+coronal] (e.g., Arabic OCP has no blockers). In order for this to happen, the training data should have sufficient evidence that co-occurrences between different coronal subcategories are underattested. As was shown in (25), no combination of the three coronal subcategories is underattested.

Even if the [+coronal] tier is discoverable or manually provided, the projection of a [+coronal] tier is in fact too general for the Arabic OCP pattern. As shown in Table 40, the three subcategories of coronals have different sets of transparent segments. Aksënova et al. (2020) point out that multiple patterns must have identical sets of irrelevant segments in order fit on a single tier. The three coronal subcategories of Arabic do not meet the above condition, which makes the [+coronal] tier unsustainable.

	stops	fricatives	sonorants
Interacting	{t d tˤ dˤ}	{s z sˤ zˤ ʃ}	{n l r}
Blocking	{}	{}	{}
Transparent	{b f m s z sˤ zˤ ʃ n l r k g q χ ʁ ħ ʕ h ʔ}	{b f m t d tˤ dˤ n l r k g q χ ʁ ħ ʕ h ʔ}	{b f m t d tˤ dˤ s z sˤ zˤ ʃ k g q χ ʁ ħ ʕ h ʔ}

Table 40. Arabic coronal subcategories

The projection of a [+consonantal] tier is also unlikely; and this in fact holds not only in Arabic but also across languages. Whatever is on the tier must participate in a pattern, either as an interacting segment or a blocker. Thus, a [+consonantal] tier would be adequate only with a consonantal interaction in which the set of relevant segment equals to the entire consonant inventory of a language; that is usually not the case. In dissimilatory consonantal interactions with no blockers, such as Arabic place OCP, constraints have a format of $*[\pm\alpha]...[\pm\alpha]$, which leads to a very specific tier $[\pm\alpha]$. Even if such restrictions also have blockers, as in Latin *l* dissimilation, blockers are usually only a subset of consonants. In assimilatory consonantal interactions, constraints might have a format of *[$\pm\alpha$]...[$\mp\alpha$] but it does not lead to projecting the entire consonant inventory on the tier because it is frequently the case that the interacting feature [α] is not specified for all the consonants. For example, in Navajo sibilant harmony harmony, the constraint *[± anterior][∓ anterior] leads to a [+strident] tier because anterior is only specified for stridents. Moreover, consonant harmony patterns are sometimes captured as an interaction between non-overlapping classes; Lamba nasal harmony can be described as *[+nasal][ ][+liquid], which leads to only a subset of the consonantal inventory of Lamba.

By contrast, there is bias towards a more general tier for vowel interactions due to the nature of vowel interactions. Since they tend to be assimilatory than dissimilatory in most cases, constraints are more likely to have a format of *[$\pm\alpha$]...[$\mp\alpha$], where the feature [α] is specified for all the vowels (e.g., [±back]), both of which together increases a chance of projecting a more general tier, such as [+syllabic].

5.3 The role of consonants in vowel interactions

My learner is based on the dichotomy between trigram-boundedness and unboundedness that can be found in phonology. The essence of my learner is the existence of the evaluation step, in which the consonant tier and the vowel tier are provided in order to determine the necessity and accuracy of candidate tiers. Under the current setup in which the learner has access only to the consonant and vowel tiers during evaluation, the learner cannot handle cases in which consonants play a role in vowel interactions or/and vice versa. For example, if round vowel harmony can be blocked by an intervening labial consonant, my learner will not be able to discover such a consonant as a blocker since the evaluation of the round vowel harmony is executed on the pre-given vowel tier, which excludes labial consonants.

The current version of my learner cannot handle such cases; but I believe that learning of such patterns should also be based on the typology of dependencies between vowels and consonants. To my knowledge, there are no cases of consonant harmony where intervening vowels play a role (Rose and Walker 2011). However, it has been attested that consonants often participate in vowel harmony via blocking or triggering the process. Thus, the learner should ideally be able to handle such patterns.

Blocker consonants are attested in vowel harmony. The most commonly attested types of consonant blockers in vowel harmony are, as intuited, the ones that are phonetically antagonistic to the harmonizing feature; the harmonizing feature simply cannot permeate through a consonant bearing the opposite feature. The classic example is Turkish (Clements et al. 1982), where a stem-final palatalized /l^j/ blocks progressive backness harmony onto suffixes and clitics, e.g. /petrol^j-dI/ → [petrol^j-dy] 'it was petrol', /usul^j-sIz/ → [usul^j-syz] 'without a system'. Although not as intuitive as antagonistic consonant blockers, there are also attested cases of vowel harmony in which sympathetic consonants, which are thought to be compatible with the harmony feature, block the process. For example, palatal harmony in Mina (Frajzyngier et al. 2005), in which front vowels trigger fronting of subsequent back vowels, is blocked if a palatal glide /j/ intervenes between the trigger and target vowels. Similarly in Nawuri, labial consonants /p b f m/ as well as the labialized velars /kp gb/ can block round harmony.

In addition to above cases where consonants interrupt vowel harmony, consonants can also initiate vowel harmony. The majority of antagonistic blockers described above actually belongs to this category; these antagonistic consonants not only stop the spread of the harmonizing feature but also start the spread of the opposite feature in the new harmony domain. For example, palatalized /l^j/ in Turkish blocks backness harmony as mentioned above but also starts front harmony onwards. In Nawuri, the glide /w/ and non-labial labialized consonants such as k^w and s^w can trigger the exact same regressive rounding harmony onto a prefix vowel as rounded vowels do.

While I focused only on consonantal interactions and vowel interactions in the dissertation, future research will hopefully expand the scope of this learner so that it is able to capture vowel patterns where consonants play a role. One possible expansion that can be made is, instead of presupposing a consonant and a vowel tier, inducing evaluation tiers based on some other metric, such as phonetic similarity. What the above cases have in common is that the consonants that are involved in vowel harmony, more specifically sympathetic blockers and harmony triggers, are the ones that share the feature that spreads in vowel harmony.

5.4 Bounded patterns with complex syllable structures

As mentioned in the earlier chapter, the evaluation method above relies heavily on the assumption that syllable structures are simple (e.g., CV). If a language exhibits a non-iterative vowel harmony and also allows consonant clusters (e.g., C**V**C.C**V**), my learner will not be able to capture such datapoints. In the evaluation phase, my learner will not proceed to a tier projection since non-iterative vowel harmony does not hold across another vowel. Thus, the final grammar will include only the local trigram constraint but not the tier version of it, mischaracterizing words with consonant clusters. Consider the case of the central dialect of Crimean Tatar in (41), taken from McCollum and Kavitskaya (2018).

(41) Evaluation of the Crimean Tatar non-iterative vowel harmony

V tier	y y i	u u ɯ	a u u	y y i
Baseline (all segs)	kyzlygi	burunɯ	asuvlu	tʃykyndir

Similar to the Kazakh vowel harmony, the central dialect of Crimean Tatar exhibits non-iterative rounding harmony that can be initiated by any non-initial position (McCollum and Kavitskaya 2018). As can be seen in (41), the forms [kyz-lyg-i] (*[kyz-lyg-y], *[kyz-lig-i]) 'autumn-NMZR-POSS.3S', [burun-ɯ] (*[burun-u]) 'nose-POSS.3S', [as-uv-lu] 'hang-GER-ADJ', and [tʃykyndir] (*[tʃykyndyr]) 'beets' are only harmonized up to two vowels, regardless of the morphological affiliation of the target or the position of the harmony trigger; the nominalizer suffix vowel is targeted in [kyz-l**y**g-i], root-internal vowels is targeted in [bur**u**n-ɯ] and [tʃyk**y**ndir], and a non-initial vowel is the trigger in [as-**u**v-lu]. In the tier-free search, in which local n-grams are found (up to $n = 3$), the learner may discover *[+round][ ][−round], if the learner sees enough evidence that sequences of [+round] and [−round] are underattested over a consonant (e.g., VCV) as in [b**u**r**u**nɯ] and [tʃ**y**k**y**ndir]. This is not entirely impossible in this language because even in a language that deviates from strict CV alternation, simple trigrams such as VCV and CVC (compared to VCCV and CVVC) will be more frequent, as Gouskova and Gallagher (2020) pointed out, which facilitates the discovery of a trigram constraint in the baseline grammar search. However, even if this constraint is successfully learned, evaluation will rule it out from a tier projection since the restriction does not hold over another vowel. At most, this restriction can be discovered again as a local trigram in the final grammar search, which is not sufficient to capture the full range of the vowel distributions because this language allows consonant clusters such as *zl*, over which the vowel harmony still holds (*[kyz-lig-i]); local trigrams are too short to penalize these forms.

In Hayes and Wilson (2008), positing a vowel tier was initially motivated by the fact that the vowel restrictions in Shona hold over a consonant cluster of any length. However, a tier projection is not the solution for non-iterative cases; if a full vocalic tier is given and a harmony-triggering constraint, such as *[+round][−round] is posited on it, then there is no way to delimit the harmony domain to a bounded window; the grammar enforces fully-iterative harmony. Shown in (42) is a tier-based grammar's prediction on languages with non-iterative harmony, fully-iterative harmony, and no harmony. As can be seen, the category of fully-iterative forms is preferred over non-iterative and no harmony languages overall. Put differently, bigram constraints on the vowel tier will collectively prefer fully-iterative forms over non-iterative one, failing to predict non-iterative harmony.

(42) Tier-based bigram constraint predicts fully-iterative harmony

		Vocalic tier *[+round][−round] $w = 1$	$\mathscr{H}$
non-iterative	kyzlygi	-1	-1
	burunɯ	-1	-1
	asuvlu	0	0
	tʃykyndir	-1	-1
fully iterative	*kyzlygy	0	0
	*burunu	0	0
	asuvlu	0	0
	*tʃykyndyr	0	0
no harmony	*kyzligi	-1	-1
	*burɯnɯ	-1	-1
	*asuvlɯ	-1	-1
	*tʃykindir	-1	-1

McCollum and Kavitskaya (2022) point out the similar issue of such a harmony-triggering constraint preferring full iterativity. Their analysis is formalized within the framework of Optimality Theory, meaning that they employ both markedness and faithfulness constraints. Shown below is their analysis of Crimean Tatar using Spread as a harmony-triggering constraint and Ident(round) as a faithfulness constraint.

(43) Non-iterative candidate is collectively harmonically bounded

	/kyz-lig-i/	Spread(round)	Ident(round)
non-iterative	kyzlygi	*	*
fully iterative	*kyzlygy		**
no harmony	*kyzligi	**	

In the tableau shown in (43), non-iterative candidates are collectively harmonically bounded. If Spread(round) outranks the faithfulness constraint, iterative harmony wins. Conversely, if the faithfulness constraint is ranked above Spread(round), there is no harmony at all. There is no ranking between the two constraints that can generate non-iterative harmony.

Thus far, I have shown that neither a tier-based bigram nor a local trigram can sufficiently capture the non-iterative harmony of Crimean Tatar, a language that allows consonant clusters. A tier-based bigram will lead to predicting a pattern of fully-iterative harmony, as demonstrated in (42). A local trigram is not long enough to penalize forms which do not conform to harmony over a consonant cluster, such as *[kyzligi]. Similarly, if a transvocalic consonantal dependency can be exhibited over a sequence of multiple vowels (e.g., **C**V.V**C**), learning can be stuck; a local trigram is not sufficient and a tier projection incorrectly predicts this pattern to be unbounded. Bounded patterns observed in languages that allow complex syllables should be captured by something in between. What I currently believe to work for these cases is a local trigram equipped with some regular expression notations, such as *[+round][+consonantal]$^{+}$[−round], similar to SPE style. It is essentially equivalent to having a vocalic tier representation only for the relevant harmony domain, which spans from the first round vowel through the next vowel, regardless of the number of intervening consonants. Similarly, a constraint such as *[+nasal][+syllabic]$^{+}$[+lateral] should be able to enforce transvocalic nasal harmony over a sequence of vowels (as opposed to unbounded). It is a more restricted search strategy than searching for any four-grams, any five-grams, and so forth, and also does not incorrectly generate full iterative pattern.

5.5 Tiers that do not form a natural class

In the Inductive Projection Learner (Gouskova and Gallagher 2020) and my learner, what is projected is the smallest superset natural class that includes the segments that *matter*. In the Inductive Projection Learner, the tier for projection is searched based on both sides of a baseline placeholder trigram; for example, a stop tier was projected based on a baseline constraint *[stop][ ][ejective] in their Quechua case study. My learner has an extra step where the boundedness of the restriction is examined and blockers of the restriction are discovered. If the restriction actually generalizes unboundedly, the smallest natural class that includes the interacting classes as well as all the blockers is projected; for example, the [−cons, −nasal] tier was projected based on the evaluation results that *[l][ ][l] is unbounded and has the blockers [b, k] in the Latin simulation.

The smallest natural class approach used in these two learners predicts that the set of segments on each tier equals to a certain natural class of the language and these are the only type of tiers that are actually learnable. However, some have argued for tiers that do not form a natural class. McMullin (2016) shows two real-life language cases. For example, to properly capture Kinyarwanda sibilant harmony, there should be a tier with all the coronal consonants of the language, except for ɽ̡: [t d s z t͡s n͡z ʂ̡ ʐ̡ ɳ͡ʐ̡ ɳ͡ɖ̡ n j ɲ]. Any attempts to describe this set of segments as a single natural class will either erroneously include ɽ̡, as in coronal consonants, or exclude consonants that need to be on the tier, as in coronal obstruents or coronal non-continuants. From the perspective of phonological rules, Mielke (2004) surveys a total of nearly 17,000 sound patterns in 561 different languages. He finds 6,077 phonologically active sets of segments in these sound patterns, of which more than one quarter cannot be described as a conjunction of features in any of the feature theories.

These typological observations are not compatible with the predictions of the current tier-inducing algorithms. This mismatch can be resolved in two different ways. First, we can posit that non-natural class tiers are learnable. There is experimental evidence that human learners are actually capable of detecting phonotactic regularities among sets of segments that do not form a natural class (Koo and Oh 2013).

Instead of positing that non-natural class tiers are learnable by humans, we can alternatively make the theory of phonological feature more expressive in a way that allows describing a non-natural class as a single natural class. The typological survey of phonological patterns in Mielke (2004) suggests that employing the set operations of union and subtraction increases the empirical coverage of a feature theory. McMullin (2016) shows specific examples where referring to the set operations of union and subtraction in describing natural classes is necessary. For example, he describes the 'the non-rhotic coronal consonants' tier that Kinyarwanda requires as a union of several natural classes, such as {sibilants} ∪ {coronal stops} ∪ {palatal consonants}, or {coronal consonants} − {ɽ}. In fact, the use of set operations in defining constraints also has been already introduced. Hayes and Wilson (2008) allow exactly one of the natural classes that form a constraint to be modified by the complementation operator (∧), to refer to any segment that is not a member of that natural class. The complementation operator can make describing a set of segments more efficient, especially a set of segments that do not form a single natural class yet behave similarly. For example, their learner captures Shona height harmony using a single constraint, *[∧−high,−low][−high,−low], instead of using two constraints as the Inductive Projection Learner and my learner do, like *[+high][−high,−low] and *[+low][−high,−low]. In addition, the Latin blockers have been often defined as "non-coronal consonants and r" in some of the literature. Incorporating such a set operation can make the description of the Latin blockers feasible.

To summarize, computational learning models should be able to predict phonological patternings of such "crazy" classes by either revisiting the assumption about the learnability of non-natural class tiers or by enriching the phonological feature theory. On the other hand, there must be some mechanism to limit the expressive power of phonological tiers. Mielke (2004), while showing that more than one quarter of his survey languages engage non-natural classes, most of the languages still show a preference towards natural classes. Although Koo and Oh (2013) show that humans can learn phonotactic restrictions defined on a non-natural class tier, the majority of experimental results suggests that humans prefer phonological patternings that can be described as natural classes (Pycha et al. 2003, Wilson 2006).

5.6 Summary

In this chapter, I provided in-depth discussions on the predictions and the assumptions that my learner makes. In §5.1, I pointed out how there needed to be some extra mechanism for my learner to discover trigram-bound patterns. I attributed the typological rarity of these patterns to the inherent learning difficulty of such patterns. In §5.2, I discussed how various factors, such as the size of natural class and the type of interaction (assimilatory *vs.* dissimilatory), contribute to the specificity of tiers. There are several assumptions that my learner was built upon; first, my algorithm presumes that the consonantal tier and the vowel tier are provided in the phase of evaluation. This prevents the learner from projecting a tier based on interactions between consonants and vowels. I suggested an alternative based on the typology of interactions between consonants and vowels; projecting an evaluation tier based on some other metric such as phonetic similarity, in §5.3. My learner is motivated by the typological observation about the robust binary distinction between trigram-bound and unbounded patterns. My algorithm relies on the universal preference toward simple syllable structures, capturing bounded patterns only with local trigram constraints. I discussed in §5.4 how the tier-based bigram approach cannot represent bounded patterns in phonology. Lastly, in §5.5, I laid out possible ways to capture phonological patterns that engage "crazy" classes (Bach and Harms 1972): either revisiting the phonological feature theory or positing that non-natural classes are learnable. While these will allow us to represent such patterns, there should also be a mechanism to characterize the human bias toward natural classes.

CHAPTER 6

CONCLUSIONS

This dissertation proposes the Restrictive Tier Learner, which automatically induces only the tiers that are absolutely necessary in capturing phonological long-distance dependencies. I exploit a typological observation as a heuristic to aid in determining the necessity and accuracy of the candidate tiers: whether the pattern holds over another segment or a subset of segments. The main findings are summarized in examples throughout this chapter.

(44) The symmetry in the typology of consonant interactions and vowel interactions allows for using a unified approach in tier induction: presupposing the consonantal and the vocalic tiers.

(45) The role of trigrams is pivotal in phonotactic learning.

I first provided an in-depth description of the typology in Chapter 2, laying groundwork for the later chapters. Specifically, I highlighted the mirrored typology of consonant interactions and vowel interactions, namely the dichotomy between trigram-bound and unbounded patterns. As summarized in (44), this observation is an important building block of my learner because it allows for a unified approach to be used for both consonant and vowel interactions. Another important piece of information is that only unboundedness implies trigram-boundedness, and not vice versa. These typological observations together shed light on the critical role of trigrams in computational learning, as summarized in (45). The premise that there is no other distance at which a restriction holds than these two lets us safely assume that searching only up to only trigrams might actually be a near-exhaustive search for local interactions. On top of that, the fact that interaction beyond a trigram window, which we need tiers for, always implies interaction within a trigram window guarantees that all necessary tiers can be discovered by looking at trigram constraints. Hence, a learner can confidently search up to trigrams for local interactions and expand its search for non-local ones from the discovered trigrams.

Chapter 3 introduced the Restrictive Tier Learner, the learning algorithm that I proposed in the dissertation. The points made in (46)−(47) are the key diagnostics used in my learner, which are inspired by the typological observations laid out in (44) and (45).

(46) The necessity of a candidate tier can be determined by reweighting the placeholder trigram on the evaluation tier.

(47) Similarly, blockers of a pattern can be discovered by reweighting a set of trigrams.

To start off the chapter, I first discussed the Inductive Projection Learner (Gouskova and Gallagher 2020), which is the base learning algorithm of my learner. I also laid out the predictions the Inductive Projection Learner makes; importantly, I showed that their learner projects an unnecessary tier when the pattern is merely trigram-bound and cannot discover blockers, leading the grammar to mischaracterize words with opaque segments (blockers). Subsequently, I illustrated the structure of the Restrictive Tier Learner in-depth, along with providing relevant real-life language examples. The core of my learner is the addition of an extra evaluation step to the existing Inductive Projection Learner (Gouskova and Gallagher 2020). In the evaluation phase, the tiers cued by baseline trigrams are examined in terms of their necessity and accuracy, using the above diagnostics. The necessity of the tier can be examined by reweighting the baseline placeholder trigram itself on the temporary tier. The accuracy of the tier can be examined by a set of trigrams, whose middle placeholder is replaced by every segment that is visible on the evaluation tier. By exploiting a typological observation as a heuristic, the learner can successfully project tiers more restrictively and accurately.

In Chapter 4, I presented five case studies to test the abilities of the Restrictive Tier Learner in capturing various long-distance dependencies that are attested in natural languages. As summarized in (48), the current version of the learner does not deteriorate what the previous learning algorithms do while showing improved performance on most cases.

(48) Adding the evaluation step to the Inductive Projection Learner improves learning results.

Most importantly, my learner has has the ability to distinguish unbounded patterns from trigram-bound ones (49).

(49) My learner projects a tier for unbounded patterns and not for trigram-bound patterns.

In my learner, restrictions that are captured as a local trigram are not carried over for a tier projection unless they pass the unboundedness test during the evaluation phase. For example, the local trigram *[+dor][ ][+dor] in the Korean case study did not pass the unboundedness test because the restriction is merely trigram-bound. For a comparison, the learning result of the Inductive Projection Learner was presented in parallel, which resulted in a projection of the unnecessary tier and a failure in discovering a trigram constraint that is actually necessary. The projection of the unnecessary tiers distracted the learner and made the learner discover several tier-based constraints that are trivial while not discovering the necessary local trigram in the Korean case study. In the Latin simulation, automatic tier projection resulted in the projection of too many tiers, causing the learner to run out of memory and crash.

There are patterns that the previous learning algorithm of Gouskova and Gallagher can handle smoothly. For these cases, my learner did not perform any differently; the evaluation did not harm the learning process. The unbounded no decay case and the distance decay case belong here. As an unbounded no decay example, the phonotactics of Quechua laryngeals were examined. The results of the Restrictive Tier Learner showed that the addition of evaluation still allows the necessary tiers to be projected. Both the Inductive Projection Learner and my learner can only represent the attested patterns of gradual distance decay as a bi-level decay at most. The learning result of Arabic showed that my learner can successfully represent bi-level distance decay by discovering a tier-based bigram and a local trigram.

The Restrictive Tier Learner did not always discover blockers accurately (50).

(50) Discovering blockers was partially successful; further research is required.

In the Shona case, the blockers that were picked out by the learner did not perfectly match the traditional description of the pattern; but the discovered blockers were actually accurate. In the Latin case study, the learner picked out not enough blockers; the learner did not find [r] as a blocker although it is the most frequently claimed blocker of the pattern, while discovering some, but not all, of the non-coronal consonants. The current learner projects the smallest superset natural class that includes all the discovered blockers and the interacting segments. Thus, not picking out the exact set of blockers did not change the natural class of projection in both case studies. If non-natural class tiers are further supported by experimental evidence, the learner should be able to identify the exact set of blockers and only those. More research is necessary to achieve this, but the partial success shown in this chapter is a good first step.

In Chapter 5, I provided in-depth discussions on the predictions and the assumptions that my learner makes.

(51) Trigram-bound patterns are hard to learn and typologically rare.

In learning simulations, trigram-bound patterns were always hard to learn; some extra mechanisms were always needed to cater to the boundedness of the pattern. Related to this observation, in §5.1, I made a connection between the learning difficulty of trigram-bound patterns and the typological rarity of such patterns, adding to the recent claims about how typology is shaped by learnability (51).

While the algorithm that I proposed in the dissertation was successful in capturing many areas of phonological dependencies, there is still more work to be done in the future. Crucially, my learner was built upon several assumptions, and hopefully those will be removed in future study.

(52) Future directions

a. Rather than stipulating consonantal and syllabic tier, evaluation tiers should also be induced, and consonant-vowel interactions should be represented.

b. Bounded patterns in languages with complex syllables should be represented.

c. Non-natural class tiers should be represented.

First, my algorithm presumes that the consonantal tier and the vowel tier are provided in the phase of evaluation. This prevents the learner from projecting a tier based on interactions between consonants and vowels. In §5.3, I suggested an alternative based on the attested patterns of interactions between consonants and vowels; inducing evaluation tiers based on other metrics such as featural similarity. And hopefully this will provide a foundation to capturing interactions between consonants and vowels (52a).

My algorithm relies on the universal preference toward simple syllable structures, capturing bounded patterns only with local trigram constraints. This approach is not sufficient if the language deviates from a simple CV alternation. Tier-based bigrams are not restrictive enough, as they incorrectly predict that the pattern holds unboundedly. This point was discussed more in depth in §5.4. In future study, there should be an appropriate mechanism to capture such patterns (52b).

Lastly, my learner can only project a tier that forms a natural class, but there are a good amount of attested patterns that engage "crazy" classes (Bach and Harms 1972). As I discussed in §5.5, either by revisiting the phonological feature theory or positing that non-natural classes are learnable, such patterns should be represented, while still restricting the theory so that the human bias toward natural classes is well characterized.

To conclude, in the dissertation, I provided a tool that captures a wider range of phonological dependencies and documented some limitations of the approach taken in my algorithm. I also sketched possible extensions that future study can pursue.

BIBLIOGRAPHY

Aksënova, Alëna, Jonathan Rawski, Thomas Graf and Jeffrey Heinz (2020). The computational power of harmony. *Oxford Handbook of Vowel Harmony, Oxford University Press, under review* .

Albright, Adam and Bruce Hayes (2006). Modelling productivity with the gradual learning algorithm: The problem of accidentally exceptionless generalizations .

Allen, Blake and Michael Becker (2015). Learning alternations from surface forms with sublexical phonology. *Unpublished manuscript, University of British Columbia and Stony Brook University. Available as lingbuzz/002503* .

Archangeli, Diana B and Douglas George Pulleyblank (1994). *Grounded phonology*, vol. 25. MIT Press.

Bach, Emmon and Robert T Harms (1972). How do languages get crazy rules. *Linguistic change and generative theory* **1**. 21.

Balakaev, Maulen B, Nikolaj A Baskakov and Smet K Kenesbaev (1962). Sovremennyj kazaxskij jazyk. *Almaty, Akademiia nauk Kazahskij SSR* .

Beckman, Jill N (1997). Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* 1–46.

Bennett, William G (2013). *Dissimilation, consonant harmony, and surface correspondence*. Ph.D. dissertation, Rutgers University Rutgers, NJ.

Berkley, Deborah Milam (2000). *Gradient obligatory contour principle effects*. Ph.D. dissertation, Northwestern University.

Berkson, Kelly Harper (2010). *The nature of optional sibilant harmony in navajo*. Ph.D. dissertation, University of Kansas.

Canavan, Alexandra, George Zipperlen and David Graff (1997). Callhome egyptian arabic speech. *Linguistic Data Consortium* .

Chomsky, Noam and Morris Halle (1968). The sound pattern of english. .

Clements, George N, Engin Sezer et al. (1982). Vowel and consonant disharmony in turkish. *The structure of phonological representations* **2**. 213–255.

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**(2). 289–337.

Cole, Jennifer and Charles Kisseberth (1994). An optimal domains theory of harmony. *Studies in the Linguistic Sciences* **24**(2). 1–13.

Cser, András (2007). On the phonotactics of liquids in Latin. In *15th manchester phonology meeting*.

Cser, András (2010). The-alis/aris-allomorphy revisited. In *Variation and change in morphology: selected papers from the 13th international morphology meeting, vienna*. 33–51.

Cser, András (2020). The phonology of classical latin. *Transactions of the Philological Society* **118**. 1–218.

Dai, Huteng (2021). Learning nonlocal phonotactics in a strictly piecewise probabilistic phonotactic model. In *Proceedings of the annual meetings on phonology*.

Dressler, Wolfgang (1971). An alleged case of nonchronological rule insertion: Flōrālis. *Linguistic Inquiry* 597–599.

Finley, Sara (2011). The privileged status of locality in consonant harmony. *Journal of memory and language* **65**(1). 74–83.

Finley, Sara (2012). Testing the limits of long-distance learning: Learning beyond a three-segment window. *Cognitive science* **36**(4). 740–756.

Frajzyngier, Zygmunt, Eric Johnston and Adrian C Edwards (2005). A grammar of mina [mouton grammar library 36]. *Berlin: Mouton de Gruyter. doi* **10**(9783110893908).

Frisch, Stefan A, Janet B Pierrehumbert and Michael B Broe (2004). Similarity avoidance and the ocp. *Natural Language & Linguistic Theory* **22**(1). 179–228.

Frisch, Stefan A and Bushra Adnan Zawaydeh (2001). The psychological reality of ocp-place in arabic. *Language* 91–106.

Gallagher, Gillian (2016). Asymmetries in the representation of categorical phonotactics. *Language* **92**(3). 557–590.

Gallagher, Gillian (2020). Strident harmony from the perspective of an inductive learner. *Phonological Data and Analysis* **2**(8). 1–29.

Gallagher, Gillian, Maria Gouskova and Gladys Camacho Rios (2019). Phonotactic restrictions and morphology in Aymara. *Glossa* .

Gouskova, Maria and Gillian Gallagher (2020). Inducing nonlocal constraints from baseline phonotactics. *Natural Language & Linguistic Theory* **38**(1). 77–116.

Greenberg, Joseph H (1950). The patterning of root morphemes in semitic. *Word* **6**(2). 162–181.

Hansson, Gunnar Ólafur (2010a). *Consonant harmony: Long-distance interactions in phonology*, vol. 145. Univ of California Press.

Hansson, Gunnar Ólafur (2010b). Long-distance voicing assimilation in berber: Spreading and/or agreement. In *Proceedings of the 2010 annual conference of the canadian linguistic association*.

Hayes, Bruce and Zsuzsa Cziráky Londe (2006). Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* **23**(1). 59–104.

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

Hayes, Bruce and Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic inquiry* **39**(3). 379–440.

Hayward, Richard J (1982). Notes on the Koyra language. *Afrika und Übersee: Sprachen, Kulturen* **65**(2). 211–268.

Hayward, Richard J (1990). Notes on the aari language. *Omotic language studies* 425–493.

Heinz, Jeffrey (2010). Learning long-distance phonotactics. *Linguistic Inquiry* **41**(4). 623–661.

Heinz, Jeffrey, Chetan Rawal and Herbert G Tanner (2011). Tier-based strictly local constraints for phonology. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies: short papers-volume 2*. Association for Computational Linguistics. 58–64.

Ito, Chiyuki (2007). Morpheme structure and co-occurrence restrictions in Korean monosyllabic stems. *Studies in Phonetics, Phonology and Morphology* **13**(3). 373–394.

Itô, Junko and Ralf-Armin Mester (1986). The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic inquiry* 49–73.

Jardine, Adam (2016). Learning tiers for long-distance phonotactics. In *Proceedings of the 6th conference on generative approaches to language acquisition north america (galana 2015)*. 60–72.

Jardine, Adam and Kevin McMullin (2017). Efficient learning of tier-based strictly k-local languages. In *International conference on language and automata theory and applications*. Springer. 64–76.

Kang, Hijo (2015). Interaction of perception and memory in segmental OCP. *Journal of the Altaic Society Of Korea* (25). 145–165.

Kim, Kyungil (1985). *A statistical analysis of the syllable structure in korean*. MA thesis, Seoul National University.

Koo, Hahn and Young-il Oh (2013). Beyond tier-based bigrams: an artificial grammar learning study. *Language Sciences* **38**. 53–58.

Martin, Andrew Thomas (2005). *The effects of distance on lexical bias: sibilant harmony in navajo compounds*. Ph.D. dissertation, Citeseer.

McCarthy, John J (1986). Ocp effects: Gemination and antigemination. *Linguistic inquiry* 207–263.

McCarthy, John J (1988). Feature geometry and dependency: A review. *Phonetica* **45**(2-4). 84–108.

McCarthy, John J (1994). The phonetics and phonology of semitic pharyngeals. *Papers in laboratory phonology III: Phonological structure and phonetic form* 191–233.

McCollum, Adam (2019). Epiphenomenal and true non-iterative harmony. *Oxford handbook of vowel harmony* .

McCollum, Adam and Darya Kavitskaya (2022). On the status of non-iterativity in feature spreading. *Glossa* .

McCollum, Adam G (2018). Vowel dispersion and Kazakh labial harmony. *Phonology* **35**(2). 287.

McCollum, Adam G and Darya Kavitskaya (2018). Non-iterative vowel harmony in Crimean Tatar. In *Proceedings of the 35th west coast conference on formal linguistics (wccfl)*. 259–268.

McMullin, Kevin and Gunnar Ólafur Hansson (2019). Inductive learning of locality relations in segmental phonology. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* **10**(1).

McMullin, Kevin James (2016). *Tier-based locality in long-distance phonotactics: learnability and typology*. Ph.D. dissertation, University of British Columbia.

Mielke, Jeff (2004). *The emergence of distinctive features*. Ph.D. dissertation, The Ohio State University.

Odden, David (1994). Adjacency parameters in phonology. *Language* 289–330.

Padgett, Jaye Edward (1991). Stricture in feature geometry .

Park, Nayoung (2020). *Stochastic learning and well-formedness judgement in korean phonotactics*. Ph.D. dissertation, Seoul National University.

Pycha, Anne, Pawel Nowak, Eurie Shin and Ryan Shosted (2003). Phonological rule-learning and its implications for a theory of vowel harmony. In *Wccfl*. vol. 22, 423–435.

Ribeiro, Eduardo Rivail (2002). Directionality in vowel harmony: the case of karajá (macro-jê). In *Annual meeting of the berkeley linguistics society*. vol. 28, 475–485.

Rose, Sharon and Rachel Walker (2011). Harmony systems. *The handbook of phonological theory* **2**. 240–290.

Sapir, Edward and Harry Hoijer (1967). *The phonology and morphology of the navaho language: University of california publications in linguistics 40*. Berkeley: University of California Press.

Shibatani, Masayoshi (1990). *The languages of japan*. Cambridge University Press.

Stanton, Juliet (2016a). Latin -alis/aris and segmental blocking in dissimilation.

Stanton, Juliet (2016b). Learnability shapes typology: the case of the midpoint pathology. *Language* **92**(4). 753–791.

Steriade, Donca (1987). Redundant values. Bosch, Anna, Barbara Need & Eric Schiller. In *23rd regional meeting of the chicago linguistics society, part*. vol. 2.

Suzuki, Keiichiro (1998). *A typological investigation of dissimilation*. Ph.D. dissertation, The University of Arizona.

Walker, Rachel (2000). Yaka nasal harmony: Spreading or segmental correspondence? In *Annual meeting of the berkeley linguistics society*. vol. 26, 321–332.

Walker, Rachel, Fidèle Mpiranya et al. (2006). On triggers and opacity in coronal harmony. In *Bls*. Citeseer. vol. 31, 383–394.

Welmers, Wm E (1973). *African language structures*. University of California Press.

Wilson, Colin (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive science* **30**(5). 945–982.

Zymet, Jesse (2014). Distance-based decay in long-distance phonological processes. In *The proceedings of the 32nd west coast conference on formal linguistics*. 72–81.