Tier-Based Locality in Long-Distance Phonotactics: Learnability and Typology

by

Kevin James McMullin

B.A., The University of North Carolina at Chapel Hill, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Linguistics)

The University of British Columbia
(Vancouver)

February 2016

Abstract

An important property of any language's sound system is its phonotactics—the unique way in which it allows its inventory of speech sounds to combine. Interestingly, certain types of phonotactic co-occurrence restrictions found in natural languages may hold across any amount of intervening material. For example, the Samala (Chumash) language of Southern California exhibits a pattern of sibilant harmony, such that [s] and [ʃ] may not co-occur anywhere within the same word (e.g. /ha-**s**-xintila-wa**ʃ**/ becomes [ha-**ʃ**-xintila-wa**ʃ**] 'his former gentile name'; Applegate, 1972).

Long-distance dependencies like this, despite being relatively common cross-linguistically, are known to pose serious problems for learnability. A learner needs an enormous amount of computational power to discover an interaction in an unbounded search space defined by arbitrary distances, resulting in patterns that are not learnable in practice. Their existence in natural languages thus suggests that humans are equipped with cognitive learning biases that restrict the available hypothesis space and facilitate the learning of patterns with certain properties but not others.

This dissertation presents a series of artificial language learning studies that support the hypothesis that the typology of locality relations in long-distance consonantal phonotactics is shaped, at least in part, by such biases. From a theoretical perspective, the goal is to explore and define the boundaries of the human learner's hypothesis space for phonotactic patterns. I argue that the seemingly simple constraints used in the Agreement by Correspondence framework (Rose and Walker, 2004; Hansson, 2010a; Bennett, 2013) generate many pathological patterns that are unattested cross-linguistically. By contrast, the properties of locality observed

for patterns of long-distance consonant agreement and disagreement belong to a well-defined and relatively simple class of subregular formal languages (stringsets) called the Tier-based Strictly 2-Local languages (TSL_2; Heinz et al., 2011). I therefore argue that class of TSL_2 stringsets offers an excellent approximation of the boundaries of possible, human-learnable phonotactics. More generally, I suggest that the formal-language-theoretic approach can be used to inform phonological theory, allowing for a better understanding of the computational complexity and learnability of predicted patterns.

Preface

All of the experimental work presented henceforth was conducted in the Language and Learning Laboratory at the University of British Columbia. All experiments and associated methods were approved by the University of British Columbia's Research Ethics Board [certificate #H13-00857].

Portions of the text related to Experiment 1 (Section 2.2) and the discussion of modular learning (Section 4.1) have been modified from published material that described preliminary results of Experiment 1 [McMullin, K. and Hansson, G. Ó. (2014). Locality in long-distance phonotactics: evidence for modular learning. In Iyer, J. and Kusmer, L., editors, *Proceedings of the 44th meeting of the North Eastern Linguistic Society*, volume 2, pages 1–14. GLSA Publications, University of Massachusetts, Amherst, MA.]. I was the lead investigator for this project, responsible for all major areas of concept formation, experimental design, stimulus preparation, data collection and analysis, as well as manuscript composition. Gunnar Ólafur Hansson was involved throughout the project, in concept formation, experimental design, data analysis, and manuscript composition.

Portions of the text related to the Agreement by Correspondence framework and the predictions thereof (Sections 2.3 and 3.1), and the Tier-based Strictly 2-Local characterization of long-distance phonotactics (Section 4.2) have been modified from an unpublished manuscript, which is based on a paper presented at the 2014 Annual Meeting on Phonology [McMullin, K. and Hansson, G. Ó. (2015). Long-distance phonotactics as Tier-based Strictly 2-Local languages. Unpublished ms. University of British Columbia.]. I was the lead investigator, and both authors contributed to concept formation, data analysis, and manuscript composition.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I am so grateful for so many people who encouraged, helped, inspired, and motivated me as I completed this task—it was not easy enough to be done on my own.

I will first acknowledge that my wife Ashley has provided me with a ridiculous amount of support during the last few years. While I can recognize certain sacrifices she made and the amount of faith she put in my abilities, I do not believe that I can fully understand how difficult it is to be on the other side of life as a Ph.D. student. Nonetheless, I offer her an assuredly heartfelt "Thank you!", along with the hope that our adventure in Vancouver was both fun and worthwhile. I love you.

I owe a great deal of thanks to my parents Stephen and Rosalie, and to my sister Nancy, who have always supported me and have rarely questioned my lofty goals. Furthermore, they continue to be excellent role models. Since I began as an undergraduate, all three of them have obtained a mid-career Masters or Ph.D. while retaining a full-time position. This astounds me, and it was often an unspoken reminder that my job was comparatively rather straightforward.

Like many others in this field, I was initially drawn to linguistics by the enthusiasm of those teaching my introductory courses. I therefore want to recognize two instructors from my time at UNC: Melissa Frazier, who was perhaps the best teacher I have ever had (despite writing her dissertation at the time) and Jen Smith, who developed my interest in phonology, and who provided me with much-needed advice and supervision from afar as I applied to graduate programs.

While at UBC, I benefited greatly from intellectual discussions and social endeavours with nearly all of my fellow graduate students. I extend my thanks to all of them, while recognizing but a subset by name: James Crippen, Ella Fund-Reznicek, Joash Gambarage, Michael McAuliffe, Stacey Menzies, Lauren Quinn, Jennifer Abel, Alexis Black, Masaki Noguchi, Sonja Thoma, Blake Allen, Andrei Anghelescu, Natalie Weber, Michael Schwan, Erin Guntly, Adriana Osa-Gómez, Emily Sadlier-Brown, Oksana Tkachman, Avery Ozburn, and Megan Keough.

I had the pleasure of working closely with a number of UBC faculty members who have influenced my research and my perception of what it means to be a scholar. In addition to those who taught my graduate courses, I would like to mention Kathleen Currie Hall and Molly Babel, who provided excellent demonstrations of what it will take to succeed early in my career.

As the non-phonologist on my committee, Carla Hudson Kam's contributions were practical in nature, and essential to my success. She provided the facilities and equipment for all of my experiments, she stressed the importance of methodological detail and statistical rigour in my research, and her role in meetings was often to ensure that my tasks could actually be achieved on time. I am left wondering how anyone completes a Ph.D. without Carla on their committee.

It is difficult to overstate the impression that Doug Pulleyblank has made on me. I have observed and interacted with him in his role as a teacher, as a researcher, as Graduate Advisor, as Department Head, and as a member of my committee. I continue to be struck by his respect for others, his leadership, his general curiosity, his brilliance, and his focus on the importance of everyone's work-life balance.

Inevitably, however, the person whose influence is most apparent in my work is Gunnar Hansson. I could not have asked for a better supervisor and I would not have been successful without him. Despite spending hundreds of hours together discussing research, I cannot pinpoint why exactly we work so well together. I can only hope that I will be able to exemplify similar qualities as a person and a linguist.

I have also benefited from the input and advice of several other people. In particular, discussion and correspondence with Will Bennett, Jeff Heinz, Adam Jardine, and Rachel Walker have provided more insight than they may realize, and my research is significantly better because of them. Furthermore, Joe Stemberger and Joel Friedman (University Examiners), as well as Sharon Rose (External Examiner) offered several suggestions that have improved my dissertation.

Finally, I point out that my dissertation research was supported financially by a SSHRC Doctoral Fellowship, a UBC Faculty of Arts Graduate Research Award, and SSHRC Insight Grant 435–2013–0455 to Gunnar Ólafur Hansson.

Chapter 1

Introduction

This dissertation explores the boundaries of the human language learner's hypothesis space with respect to the types of sound patterns that can be inferred from exposure to linguistic data. I focus mainly on phonotactic dependencies that hold between non-adjacent consonants and, more specifically, on the types of locality relations that are found in such patterns. The primary goal is to establish a computationally well-defined set of sound patterns that are predicted to be human-learnable. This is achieved by assessing the range of cross-linguistically attested patterns in light of experimental results from a series of artificial language learning studies.

As an example of a dependency that holds between non-adjacent consonants, consider the case of sibilant harmony in Samala (Ineseño Chumash; Applegate, 1972), in which two sibilants are not permitted to co-occur anywhere within the same word unless they agree in anteriority. Evidence of this can be seen when the sibilants that surface as [+anterior] segments [s] or [tsʰ] in the word [**s**api**tsʰ**olu**s**] 'he has a stroke of good luck' instead surface as [–anterior] segments when the perfective suffix [-waʃ] is added, becoming [**ʃ**api**tʃʰ**olu**ʃ**wa**ʃ**] 'he had a stroke of good luck' (Applegate, 1972, p. 119).

Patterns like this, which sometimes hold across a great number of segments or syllables (e.g. /k-su-k'ili-mekeken-ʃ/ becomes [k**ʃ**uk'ilimekeke**tʃ**] 'I straighten myself up'; Applegate, 1972, p. 119), are known to pose serious problems for learnability. This is true both from a general perspective on the cognitive limits of human learning (see, e.g., Creel et al., 2004; Newport and Aslin, 2004; Gebhart et al., 2009), as well as for computational models of learning (see, e.g., Heinz, 2007, 2010; Hayes and Wilson, 2008; Heinz et al., 2011; Goldsmith and Riggle, 2012; Jardine and Heinz, 2015). The core issue is that if sound patterns can hold across unknown, arbitrary, and potentially unbounded distances, the search space of possible patterns is simply too large for any learner (human or otherwise) to traverse efficiently. Nonetheless, long-distance dependencies are robustly attested in natural language, suggesting that some kind of cognitive *learning biases* must be present that enable successful learning of the exhibited sound patterns. The central questions guiding this dissertation are as follows:

- Does the typology of attested non-adjacent phonotactic patterns reflect the properties and inductive biases of human learning mechanisms?
- Do existing theories of long-distance consonant interactions over- or under-predict with respect to the range of patterns supported by empirical data?

Upon investigating the above questions, it becomes clear that current approaches do not necessarily offer the correct set of predictions about the typology and learnability of long-distance dependencies. This motivates a third question:

- Can a computational (formal-language-theoretic) definition of the notion of a *tier* be incorporated into phonological theory in order to improve predictions about the set of possible, human-learnable phonotactic patterns?

The present research therefore contributes to a growing body of literature that highlights the importance of a computational foundation for phonological theory, and an explanation of the relationship between the typology and the learnability of linguistic patterns (see also, e.g., Heinz, 2007, 2010; Finley, 2008; Moreton, 2008, 2012; Hayes and Wilson, 2008; Lai, 2012; Morley, 2015).

The remainder of this chapter establishes a context for asking the above questions and defines the scope of my dissertation research. Section 1.1 outlines the basic assumptions I make about the relationship between linguistic typology and language learning. Section 1.2 motivates the need for further behavioural data about the learnability of linguistic sound patterns and summarizes the experimental paradigm of artificial language learning—a methodology that has seen a growing amount of attention in the literature, and which represents a sizeable portion of the research presented in this dissertation. In Section 1.3, I introduce the empirical focus of my dissertation, long-distance consonantal phonotactics, summarizing the basic cross-linguistic properties with respect to similarity, locality relations, and blocking, as well as any existing experimental evidence of learning biases that are associated with these aspects of the typology. Section 1.4 then outlines two distinct theoretical frameworks, both Optimality Theory and formal language theory, and summarizes previous theoretical accounts of long-distance consonant interactions. Section 1.5 states the central claims that will be made in this dissertation, and Section 1.6 outlines the structure of the remaining chapters.

1.1 Typology and Learning Bias

Patterns that are observed in natural language, irrespective of how they first arise, must be learnable in order to be acquired by a new generation of speakers. Cross-linguistic typological distributions may therefore provide an indirect window on properties of the human language learner, such as restrictions on the available hypothesis space or heuristics employed to navigate that space. Such learning biases, be they domain-specific (i.e. at play only in linguistic or phonotactic learning) or domain-general (applying also when learning, for example, visual or non-linguistic auditory patterns) are a major factor in shaping and constraining typological variation (see, e.g., Wilson, 2006; Finley and Badecker, 2007; Kirby et al., 2008; Moreton, 2008, 2012; Scott-Phillips and Kirby, 2010; Culbertson, 2012; Culbertson et al., 2012; Rafferty et al., 2013).

Figure 1.1 provides a basic illustration of the connections between a learner, a learner's hypothesis space, and a set of cross-linguistically attested patterns. Consider a human learner whose goal is to correctly identify a grammar for a language L_x after being exposed to a finite amount of primary linguistic data (from L_x). If the data is drawn from a natural human language, we would certainly expect the learner to acquire the correct grammar. I assume that there are restrictions on the types of languages that the human learning mechanism is capable of identifying correctly, and one of the goals of this research is to determine exactly what that boundary is. I note, however, that in order to simplify the present investigation, I largely set

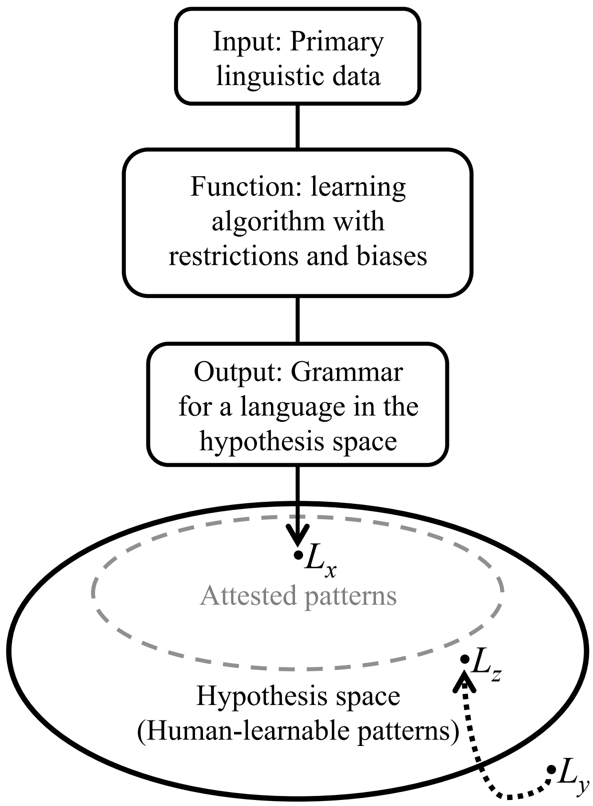

Figure 1.1: Basic model of a learning as a function that maps input training items to an output language in the range of the learner.

aside issues of imperfect learning and assume an idealized learner that is successful whenever the language belongs to the learner's hypothesis space (though see, e.g., Kiparsky, 1968 on imperfect learning and analogical change).

Imagine now that the learner is exposed to data from a hypothetical, unattested language L_y that contains a pattern that does not resemble anything found in the typology of natural language. If such a case were to arise, we might expect that the learner would erroneously map the training data from L_y to an alternative grammar that generates a language that is somewhat similar to L_y, but within range of the learner (L_z in Figure 1.1). The types of linguistic patterns that remain stable, persisting through transitions from one generation of speakers to the next, should be those that are human-learnable.

While the set of long-distance phonotactic dependencies observed in natural language may provide a rough estimate of the learner's hypothesis space, I also assume that there exist many accidental gaps in the typology—patterns that a human learner would correctly identify if provided the right training data. With respect to phonological patterns in particular, there also exist certain stochastic pressures shaping the typology in terms of which patterns are most likely to arise in or disappear from a language through, e.g., misperception, misproduction, errors in speech planning, and so on (Ohala, 1993; Blevins, 2004; Hansson, 2008; Garrett and Johnson, 2012). The present research does not attempt to capture these influences on the relative frequency and distribution of all possible patterns, but focuses on establishing a definition of the categorical boundary between possible (human-learnable) and impossible (not human-learnable) long-distance phonotactic patterns, by evaluating the predictions of multiple hypotheses within two theoretical frameworks (see Section 1.4).

1.2 Experimental Methodology

Under the assumption that all patterns exhibited in natural language are learnable, any proposed region of possible languages must contain at least those patterns that are attested. While the range of observed patterns is likely a reasonable estimate of the limits on human learning, it cannot be taken for granted that other, logically possible but unattested patterns are unlearnable. As a means of testing the limits of human learning, this dissertation employs an *artificial language learning* paradigm. This methodology has become increasingly popular for linguists and cognitive psychologists interested in language learning (e.g. Pycha et al., 2003; Wilson, 2003, 2006; Hudson Kam and Newport, 2005, 2009; Finley and Badecker, 2009; Moreton, 2008, 2012; Finley, 2011, 2012; Moreton and Pater, 2012a,b). One of the main advantages of such methods is that they enable the researcher to overcome the relative rarity (or non-existence) of certain patterns, which makes it unfeasible to use the more traditional method of studying children throughout the process of language acquisition.

In a typical experiment, subjects complete a training phase in which they are exposed to certain forms from an artificial language, constructed by the researcher, that exhibits the pattern of interest. This is followed by a testing phase to determine whether or not they have learned the pattern, and in some cases whether or not they generalize it to novel contexts that were not encountered in training. The artificial language learning methodology therefore provides an accessible way to obtain data about the learning of any type of phonological pattern, whether it is relatively frequent, rare, or completely unattested across the world's languages. An additional benefit of the paradigm is that the researcher has total control over the learner's input, such that direct comparisons between learners who received only slightly different sets of training items are possible. However, the methodology is not without criticisms (for a recent overview of the findings and criticisms of artificial phonology experiments, see Moreton and Pater, 2012a,b). For example, we may not know what biases the learners are bringing in from their own native language, or their language experience as a whole, and there has been little research concerning the relationship between artificial language learning and natural language learning (L1 or L2 acquisition; see Ettlinger et al., 2015, for a summary of the limited evidence, and an argument that artificial language learning is similar to second language learning). It is thus imperative to include a well-constructed control condition that can indicate any biases a learner might come in with or that result from the task itself, so that we can factor them out when performing statistical analyses (Reber and Perruchet, 2003; Finn and Hudson Kam, 2008).

1.3 Typology of Long-Distance Consonant Interactions

While many phonotactic patterns result from interactions between adjacent segments (e.g. voicing or place assimilation in consonant clusters, palatalization of consonants before front vowels), there are several types of phonological patterns that apply even across intervening material. Such patterns have long been a topic of debate in phonological theory (e.g. Halle and Vergnaud, 1981; Poser, 1982; Steriade, 1987a,b; Odden, 1994; Gafos, 1999; Hansson, 2001, 2010a; Ní Chiosáin and Padgett, 2001; Pulleyblank, 2002; Rose and Walker, 2004; Nevins, 2010; Bennett, 2013). In what follows, I use the term *long-distance phonotactics* to refer to co-occurrence restrictions on segments in surface forms, primarily with respect to non-adjacent pairs of *consonants* in particular. Such interactions may be assimilatory or dissimilatory in nature, but must hold between two consonants that are separated by at least an intervening vowel (Hansson, 2010a; Bennett, 2013). For example, as the data in (1) and (2) demonstrate (with evidence from suffix allomorphy), certain languages may require two co-occurring liquids [l, r] to agree (as in Bukusu; Odden, 1994), or to disagree (as in Georgian; Fallon, 1993; Odden, 1994), even when they are separated by several segments.

(1) Liquid harmony in Bukusu (Bantu; Odden, 1994)

a.	teex-e**l**-a	'cook for'
b.	**l**im-i**l**-a	'cultivate for'
c.	ka**r**-i**r**-a	'twist'
d.	**r**um-i**r**-a	'send someone'
e.	**r**eeb-e**r**-a	'ask for'

In the data from Bukusu above, (1a) and (1b) show that when a verb stem contains no liquids, or when it contains a liquid [l], the applicative suffix surfaces as [-il] or [-el]. However, when the stem contains [r] as in (1c)-(1e), there is an alternation in the suffix, which becomes [-ir] or [-er]. The resulting generalization is that, in Bukusu, words with *[r…l] are not permitted.[1]

The opposite generalization is shown by the Georgian data presented below:

(2) Liquid dissimilation in Georgian (Kartvelian; Fallon, 1993; Odden, 1994)

a.	dan-u**r**i	'Danish'
b.	p'o**l**on-u**r**i	'Polish'
c.	ung**r**-u**l**i	'Hungarian'
d.	ap**r**ik'-u**l**i	'African'

The suffix /-uri/ remains faithful in (2a)-(2b), when there is no other [r] preceding the suffix. However, if another [r] does precede it, the suffix surfaces as [-uli], as seen in (2c)-(2d). The generalization for the data in (2) is that in a word that

[1] This is the extent of the data as described by Odden (1994, based on his own field notes). Hansson (2010a) further notes that the pattern holds morpheme-internally, and that it may be optional in the longer-range contexts (e.g. [rum-ir-a] [rum-il-a] 'send for'; cited from the *Comparative Bantu Online Dictionary* available at http://www.cbold.ish-lyon.cnrs.fr/).

contains two liquids, they may not both be [r].[2]

A practical motivation for my focus on long-distance consonant interactions is the relative recency and accessibility of comprehensive typological studies, both for consonant harmony (Hansson, 2001, 2010a; Rose and Walker, 2004) and long-distance consonant dissimilation (Suzuki, 1998; Bennett, 2013). The cross-linguistic distribution of these patterns reveals several interesting asymmetries that I hypothesize are related to human learning biases, but which have not been fully investigated in artificial language learning studies. The remainder of this section summarizes the typology and relevant experimental results for three important aspects of long-distance consonant interactions: the relative similarity of interacting segments, the distance between them, and whether or not the dependency can be blocked when a specific segment intervenes.

1.3.1 Similarity

First, with respect to *similarity* in terms of shared features, it seems that there is a general dispreference for the co-occurrence of highly similar segments, which can be repaired either with harmony by making them even more similar (perhaps identical), or with dissimilation by differentiating them beyond some threshold. For example, the most common type of consonant harmony (by far; Hansson, 2010a) is sibilant harmony, which prohibits the co-occurrence of two sibilants that do not match for some other feature (e.g. anteriority, voicing, or both). Likewise, a well-attested type of dissimilation requires disagreement in major-place features among obstruents (Alderete and Frisch, 2007; Bennett, 2013). I note, however, that there is a difference between harmony and dissimilation in the types of features that are most often involved. In contrast to the relatively common patterns of sibilant harmony and major-place dissimilation, there are no attested cases of a language that prohibits *[s…s] and *[ʃ…ʃ], but allows [s…ʃ] and [ʃ…s] (i.e. a form a sibilant dissimilation), nor is there a language that exhibits a pattern of major-place harmony (though the latter is commonplace in child language; Levelt, 2011). The present research largely sets aside issues of similarity and feature specifications (for an extensive investigation of the relationship between the cross-linguistic properties of

[2]This is a simplified dataset presented for introductory purposes, as the ban on words containing *[r…r] is blocked by an intervening [l]. For a full discussion of the Georgian data, see Section 4.1.2.

long-distance consonant assimilation and dissimilation, see Bennett, 2013).

The experimental literature, linguistic and non-linguistic alike, supports the idea that human learning of a dependency is influenced by the similarity of the two elements that enter into that dependency. Evidence from a number of studies shows that non-adjacent dependencies are more easily learned when the two interacting elements are more similar (e.g. Creel et al., 2004; Newport and Aslin, 2004; Gebhart et al., 2009). I note, however, that evidence from Koo and Oh (2013) suggests that similarity among interacting consonants may be best understood as a contributing factor to learnability rather than a necessary condition for human learning (see Section 5.1.3 for a full description of their study).

1.3.2 Distance

Apart from similarity, the relative *distance* separating the two segments can also be a conditioning factor. Odden (1994) describes the Bukusu and Georgian patterns, in (1) and (2), respectively, as *unbounded* dependencies that hold across any number of non-participating interveners. He further argues that syllable adjacency (i.e. $\ldots\underline{\mathrm{C}}\mathrm{v}.\underline{\mathrm{C}}\ldots$ contexts) is the appropriate characterization of locality for certain non-adjacent phonotactic dependencies (in addition to a third level of locality which requires direct adjacency in the string). Similarly, Pulleyblank (2002) proposes a set of constraints that drive harmony and dissimilation, which are specified for three discrete levels of locality—'Distant' (unbounded), 'Medium' (roughly equivalent to syllable-adjacent), and 'Close' (string-adjacent).

The starting point of this dissertation is an extensive investigation of the distinction between unbounded patterns, which hold in all $\ldots\underline{\mathrm{C}}\ldots\underline{\mathrm{C}}\ldots$ contexts, and dependencies that apply only within a bounded $\ldots\underline{\mathrm{C}}\mathrm{v}.\underline{\mathrm{C}}\ldots$ window. For consonant harmony in particular, this is a robust dichotomy and there is no other type of restriction on distance that is attested, such as a dependency that holds across at most one intervening consonant.

An example of the dichotomy is illustrated below with cases of sibilant harmony from two related Omotic languages. The data in (3) show an unbounded dependency found in Aari. The perfective suffix /-s/, which stays faithful in (3a), surfaces instead as [ʃ] when it is preceded (at any distance) by a lamino-postalveolar sibilant, as seen in (3b)-(3d).

(3) Unbounded sibilant harmony in Aari (Hayward, 1990)

a.	/baʔ-s-e/	ba**ʔs**e	‘he brought’
b.	/ʔuʃ-s-it/	ʔu**ʃʃ**it	‘I cooked’
c.	/tʃ’a̤ːq-s-it/	**tʃ**’a̤ːq**ʃ**it	‘I swore’
d.	/ʃed-er-s-it/	**ʃ**eder**ʃ**it	‘I was seen’

In the data from Koyra below, the 3rd person masculine singlular (perfective) suffix /-osːo/ harmonizes with a preceding sibilant when they are separated by a single vowel, as in (4b)-(4c). However, it surfaces faithfully across any greater distance (i.e. when another surface consonant intervenes), as shown in (4d)-(4e).

(4) Transvocalic sibilant harmony in Koyra (Koorete; Hayward, 1982)

a.	/tim-d-osːo/	tindo**s**ːo	‘he got wet’
b.	/patʃ-d-osːo/	pa**tʃ**ːo**ʃ**ːo	‘it became less’
c.	/giːʒ-d-osːo/	giː**ʒ**ːo**ʃ**ːo	‘it suppurated’
d.	/ʃod-d-osːo/	**ʃ**odːo**s**ːo	‘he uprooted’
e.	/ʔatʃ-ut-d-osːo/	ʔa**tʃ**utːo**s**ːo	‘he (polite) reaped’

Note that while Rose and Walker (2004) and Bennett (2013) follow Odden (1994) in defining the latter of these two patterns in terms of *syllable-adjacency* (since they are most often observed in a …CV.CV… configuration), I instead follow Hansson (2010a) in characterizing them as *transvocalic* harmony since the dependencies appear to hold across maximally one vowel (short or long) and are never seen to hold across an intervening consonant. The crucial data needed to distinguish between the syllable adjacent vs. transvocalic definitions of locality would be words with at least one closed syllable, such that the two elements of the dependency are in adjacent syllables but are separated by an intervening consonant (e.g. CVC.CV or CV.CVC). Hansson (2010a) notes, however, that many of the potentially informative languages do not allow coda consonants, and those that do only permit a limited set of consonants in coda position. While this topic merits future investigation, Hansson argues that what little evidence there is favours the transvocalic (rather than syllable-adjacent) characterization of these …Cv.C… dependencies.[3]

For patterns of sibilant harmony, results from artificial language learning studies align with the observed split between transvocalic and unbounded locality. Learners who encounter sibilant harmony only in transvocalic Sv.Sv contexts (where S represents a sibilant [s] or [ʃ]) tend not to generalize to greater distances, but exposure to Sv.cv.Sv harmony, across an intervening consonant, results in the learning of a genuinely unbounded pattern—subjects tend to generalize both to lesser and greater distances. This is shown by Finley (2011, 2012), who exposed subjects to a left-to-right pattern of sibilant harmony in the form of a [-su] vs. [ʃu] suffix alternation triggered by a sibilant in the stem. McMullin and Hansson (2014, Experiment 1) replicate this finding with right-to-left directionality, as subjects were exposed to sibilant alternations in 'verb' stems triggered by two separate suffixes, [-su] and [-ʃi], which indicated 'past' and 'future' tense, respectively. Finally, Experiment 1 of this dissertation (see Section 2.2.2) produces a further replication of these results, but with a different type of interaction in terms of the segments involved (i.e. liquids rather than sibilants).

With respect to long-distance consonant dissimilation, results from Koo and Cole (2006) provide evidence that liquid dissimilation can be learned from laboratory exposure to the pattern, but to my knowledge there has been no investigation into the learnability of different locality parameters for such patterns. Experiment 2 (see Section 3.2.2) fills this gap, showing that subjects learn and generalize long-distance patterns of liquid dissimilation in the same way as they do for harmony.

1.3.3 Blocking

In terms of whether or not certain intervening segments can block a long-distance interaction between consonants, the cross-linguistic details have historically been reported as being quite different for harmony and dissimilation.

The typology of consonant harmony, as reported by Hansson (2001), as well as Rose and Walker (2004), reveals a conspicuous absence of systems that exhibit

[3]Note that in (4d)-(4e), under the standard assumption that the two halves of a geminate consonant straddle the syllable boundary, the two sibilants are technically in adjacent syllables (e.g. [ʔa.**tʃ**ut.tos.**s**o] 'he (polite) reaped'), but do not agree for anteriority. This is part of Hansson's (2010a) evidence that the dichotomy is between unbounded and transvocalic (not syllable-adjacent) dependencies.

blocking effects. More recently, however, at least three languages with a genuine case of consonant harmony with blocking have been reported in the literature, including Slovenian (Jurgec, 2011, see Section 4.2.1), Kinyarwanda (Walker and Mpiranya, 2005; Hansson, 2007; Walker et al., 2008, see Section 5.1.1), and Imdlawn Tashlhiyt (Elmedlaoui, 1995; Hansson, 2010b, see Section 5.3.1.2). By contrast, no such gap in the typology has been proposed for long-distance dissimilation (to my knowledge). This may be in part due to the case of Latin liquid dissimilation, which has been studied extensively in the literature (e.g. Watkins, 1970; Dressler, 1971; Jensen, 1974; Steriade, 1987a; Odden, 1994; Cser, 2010). The traditional description of the pattern is that underlying /-al/ surfaces as [-ar] when an [l] precedes it (e.g. /lun-al-is/ becomes [lun-ar-is] 'lunar'), but that the dependency is blocked by an intervening [r] (e.g. /flor-al-is/ surfaces faithfully as [flor-al-is], not *[flor-ar-is]).[4] Despite the relative prominence of the Latin pattern in the literature, however, there do not appear to be as many instances of long-distance consonant dissimilation with blocking as one might think. In particular, Bennett (2013) notes only three cases of long-distance dissimilation that is blocked by an intervening consonant, along with four others that are either blocked in some other way, or that are not fully supported empirically.

There is a need for an investigation of long-distance dependencies with blocking in terms of whether or not humans can learn such patterns in the laboratory, but this remains outside the scope of the present research. Instead, I will argue for a unified account of the cross-linguistic properties of locality and blocking, and the resulting predictions will serve as the basis for artificial language learning studies in future research.

1.4 Theoretical Frameworks

This dissertation considers several potential characterizations of the learner's hypothesis space, but is restricted in scope to a comparison of proposals within two theoretical frameworks in particular: Optimality Theory and formal language theory. Each presents different challenges for and predictions about learnability, which

[4] Cser (2010) argues that this simple generalization is not sufficient for capturing the full regularity of the pattern. Specifically, he presents evidence from a corpus that shows that dissimilation is also blocked by intervening labial and velar consonants. This is described in detail in Section 5.1.2.

are outlined in general terms below.

1.4.1 Optimality Theory

In *Optimality Theory* (OT; Prince and Smolensky, 2004), the learner's hypothesis space is determined by an innate, universal set of ranked and violable constraints. The types of phonotactic patterns that the learner needs to consider are restricted to those generated by the factorial typology (i.e. all possible constraint rankings). If each attested pattern can be generated with at least one such ranking, then the learning problem is quite simple—the learner needs only to find a ranking that accounts for all of the encountered forms. Finding a correct constraint ranking is a relatively straightforward task if the learner has inherent access to the constraint set, and there are several algorithms that can do so (e.g. Boersma, 1997; Tesar and Smolensky, 2000; Goldwater and Johnson, 2003). In short, the success of these learning algorithms is a result of the structural properties of OT grammars, which may be independent of the constraints themselves (Heinz, 2009; Tesar and Smolensky, 2000; Dresher, 1999 makes a similar argument about learning in Principles and Parameters frameworks).

If a new pattern is discovered that cannot be accounted for with the set of posited constraints, it is not a problem for learnability itself, as the phonologist needs only to formulate a new (though by hypothesis still innate) constraint that can account for the pattern. While there are rough criteria for what constitutes a plausible and well-formed constraint, especially in terms of phonetic grounding (Archangeli and Pulleyblank, 1994; Hayes, 1999), I argue that constraints must exhibit a further *computational* grounding. For example, the number of potential violations of gradient ALIGN constraints grows quadratically with respect to the length of the word. Aside from predicting a number of unattested patterns (Eisner, 1997; McCarthy, 2003), Riggle (2004) shows that they are formally too complex for computing optimization over. I note that this dissertation does not present any arguments against constraint-based approaches or optimization in general, but instead suggests that if the goal is to achieve a feasible model of the human grammar and learning mechanism within a constraint-based framework, we need a better understanding of the computational properties underlying individual constraints, and that they should be demonstrably learnable rather than provided a priori (see also, e.g., Ellison, 1992, 1994; Hayes and Wilson, 2008; van de Weijer, 2014).

Specifically, my dissertation focuses on the predictions of Agreement by Correspondence (ABC), a framework within OT for analyzing non-adjacent segmental interactions (Walker, 2000a,c; Hansson, 2001, 2010a; Rose and Walker, 2004). This approach has seen relative success in accounting for the typology of consonant harmony, and has more recently been extended to analyses of long-distance consonant dissimilation (Bennett, 2013, 2015). The idea, motivated by the fact that interaction seems to be facilitated primarily by similarity in terms of shared features (see Section 1.3.1 above), is that a similarity-based surface correspondence relation brings two (non-adjacent) consonants into each other's purview, and that certain restrictions may be imposed on consonants that are in correspondence. The basic ABC framework is outlined in Section 2.3, and in Chapter 3 I argue that defining the human learner's hypothesis space in terms of a factorial typology of ABC constraints does not offer a satisfactory approximation of the range of patterns indicated by the empirical evidence.

1.4.2 Formal Language Theory

I investigate issues of computational complexity and learnability of phonotactic patterns within the framework of *formal language theory*. From this perspective, languages can be thought of as sets of grammatical words, whose members include only the sequences of sounds that are well-formed in the language. A phonotactic pattern is thus manifested as a restriction on the strings of segments that are permitted in the set. Strings that do not adhere to the pattern will be ungrammatical, and are therefore not members of the stringset. As the scope of this dissertation is limited to phonotactic complexity in particular, the terms *pattern* and *language* are used interchangeably in reference to the stringset (or formal language) that reflects the phonotactics of a language.

A long-known property of phonological mappings (e.g. input strings to output strings) is that any pattern that can be generated with an ordered set of rewrite rules belongs to the class of regular relations (Johnson, 1972; Kaplan and Kay, 1994). Consequently, as Rabin and Scott (1959) show, all stringsets generated

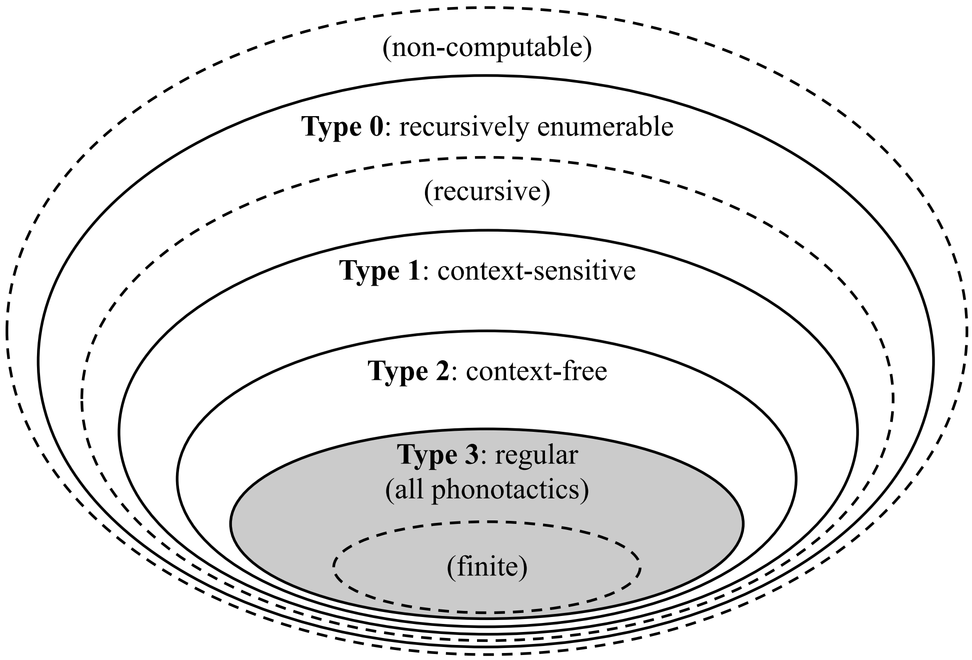

Figure 1.2: The Chomsky hierarchy. The shaded region indicates that the complexity of all attested phonotactic patterns seems to be (at most) regular.

by these relations (i.e. the surface phonotactics) are members of the regular region of the Chomsky hierarchy (Chomsky, 1956), which includes several well-known classes of formal languages that are in a subset relationship, as illustrated in Figure 1.2. Certain syntactic processes are known to result in relatively complex context-sensitive stringsets (strings of words rather than segments; e.g. Culy, 1985; Shieber, 1985; Kobele, 2006), but it turns out that all attested phonological patterns are indeed regular, including long-distance consonant agreement and disagreement (Heinz, 2010; Heinz et al., 2011; Payne, 2014). While the resulting phonotactic patterns are therefore also regular, not every pattern that can be described as a regular stringset is attested in natural language, such as a dependency that holds between the first and last segments of a word (Lai, 2012). However, the regular region can be further broken down into a hierarchy of well-studied formal language classes that are proper subsets of the regular languages (the *subregular hierarchy*; McNaughton and Papert, 1971; Rogers et al., 2010; Heinz et al., 2011; Rogers and Pullum, 2011).

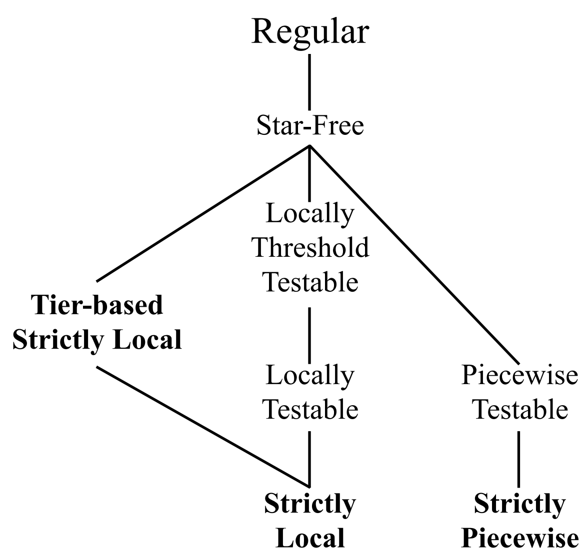

Figure 1.3: Illustration of the subregular hierarchy. The largest class of formal languages (i.e. Regular) is presented on the top, and subset classes are presented below. Each language class is thus a proper subset of any class that is above it and connected by a line. Subregular language classes that are most relevant to this dissertation are presented in boldface.

Throughout this dissertation, several of the subregular classes shown in Figure 1.3 will be assessed as potential bases for defining a boundary between possible and impossible phonotactic patterns. In other words, this boundary can be thought of as a learning bias that restricts the hypothesis space for phonotactic patterns to exactly that class of stringsets.

For instance, one type of phonotactic co-occurrence restriction regulates sequences of string-adjacent segments up to some length *k* (*k*-factors; roughly equivalent to the concept of *n*-grams). Each such restriction defines a member of the Strictly *k*-Local class of formal languages (SL_k). An example of a phonological constraint corresponding to an SL_2 language would be *[−son, αvoi][−son, −αvoi] (cf. AGREE[voice]; Lombardi, 1999), which bans any mixed-voicing sequence of two obstruents: *bk, *zt, *gθ, *pz, *xð, etc. SL_k languages thus provide sufficient expressivity for describing interactions between consonants within a particular window (bounded by *k*). If $k = 3$, we can also capture transvocalic consonant interactions as members of the SL region. For example, we might propose a constraint like *[+strid, αant][+voc][+strid, –αant] to account for transvocalic sibilant harmony, which can be thought of as an SL_3 language that disallows sequences such as *sadʒ, *tʃ'oz, *tsuʃ, etc. However, the SL region is not a plausible measure of the limits on human learning, since it is too restrictive to allow for the unbounded types of long-distance interactions described in the previous section.

By referring instead to precedence relations (i.e. $x \ldots y$), which are by definition blind to distance and intervening material, an unbounded dependency can be described as a Strictly 2-Piecewise (SP_2) pattern that disallows certain subsequences of length 2, such as *s…ʃ, *ʃ…s for patterns of sibilant harmony (Heinz, 2010; McMullin and Hansson, 2014). However, this characterization only works well for unbounded phonotactic dependencies without blocking (Heinz, 2010)—and although segmental blocking effects are relatively rare in long-distance patterns of consonant (dis)agreement, they are nonetheless attested (see Section 1.3.3) and must be accounted for.

As an alternative, unbounded sibilant harmony can be thought of as a restriction against contiguous [+strid, αant][+strid, –αant] segment pairs (*sʃ, *ʃs), where adjacency is crucially assessed only among sibilant consonants within the string. More generally, patterns that can be described in similar terms are members of the *Tier-based Strictly 2-Local* class of formal languages (TSL_2; Heinz et al., 2011). In brief, a grammar for a TSL_2 language is defined by the relevant subset of the inventory that comprises the 'tier' *T*, and the set of segment 2-factors (bigrams) that are permitted on *T*, denoted *S* (or *R* for the set of prohibited 2-factors). I will demonstrate that this characterization of long-distance dependencies, while still relatively simple, offers an account of the typological properties of locality that extends straightforwardly to cases of blocking. I note that the concept of tiers (or projections of segments) has long been used in theoretical phonology (e.g. Clements, 1980, 1985; Shaw, 1991; Odden, 1994; Blevins, 2004; Clements and Hume, 1995), but that the TSL_2 approach differs primarily in that a tier can be defined by any subset of the segment inventory (i.e. it need not be a set of segments sharing some feature, or belonging to a natural class, etc). I will argue that this formal characterization of a tier offers enough flexibility to account for a number of otherwise problematic patterns, without suffering consequences from the perspective of computational theory and learnability (Heinz et al., 2011; Jardine and Heinz, 2015).

1.5 Central Claims

The evidence presented in this dissertation points to a fundamental relationship between the cross-linguistic distribution of long-distance phonotactic dependencies and the way humans can be seen to learn those patterns in a set of artificial language learning experiments. I argue that an adequate theory of long-distance consonant interactions should account for this connection, but that the inherent properties of the Agreement by Correspondence framework (Rose and Walker, 2004; Hansson, 2010a; Bennett, 2013) result in a number of pathologies and predictions about learnability that are not borne out empirically. As such, I claim that the factorial typology of ABC constraints is not a satisfactory definition of what constitutes a possible, human-learnable phonotactic pattern. In approaching the problem from the perspective of formal language theory, the hypothesis that all long-distance phonotactic patterns are members of the Tier-based Strictly 2-Local class of stringsets (Heinz et al., 2011) withstands the scrutiny of comparison with a wide range of patterns observed in natural language. I therefore claim that the TSL_2 region of formal languages is an excellent, and moreover computationally well-defined, approximation of the learner's hypothesis space that offers an account of both unbounded and transvocalic patterns, with the latter being a special case of a more general category of long-distance dependencies with blocking. I will argue that this solution can be thought of as an independent theory of long-distance dependencies, but that it also has the potential to be integrated with constraint-based frameworks in the form of computationally grounded and learnable markedness constraints that are defined as individual formal languages (stringsets) of the TSL_2 class.

1.6 Structure of the Dissertation

The remainder of this dissertation proceeds as follows. Chapter 2 looks in more depth at the attested two-way split between unbounded and transvocalic locality for patterns of consonant harmony. After presenting the results of Experiment 1 as evidence that humans learn and generalize patterns of liquid harmony in a way that mirrors the cross-linguistic properties of locality, I offer an explanation of the dichotomy in terms of the constraints used in the Agreement by Correspondence framework. While the ABC model thus seems to provide a satisfactory account of certain empirical findings, Chapter 3 demonstrates that a number of pathological patterns can be generated within the factorial typology of ABC constraints, and argues that neither the cross-linguistic typology nor the results of Experiment 2 (an artificial language learning study of liquid dissimilation) provide support for the predictions of the ABC model. Chapter 4 turns to the alternative perspective of formal language theory, describing how phonotactic patterns can be characterized as members of several different types of subregular formal language classes (stringsets), focusing primarily on patterns that can be described as Tier-based Strictly 2-Local stringsets. After demonstrating that the cross-linguistic typology of locality relations and blocking are easily captured by varying the set of segments that constitutes the 'tier' in the TSL_2 grammar, I present experimental evidence that patterns outside of the TSL_2 region (i.e. harmony or dissimilation that applies only in 'beyond-transvocalic' contexts) are simply not learned in the laboratory (Experiments 3 and 4). Chapter 5 further scrutinizes the TSL_2 approach, arguing that the proposed region is not too big, demonstrating that it is computationally learnable, and finally considering the possibility of defining markedness constraints (i.e. co-occurrence restrictions) as individual TSL_2 languages. Finally, Chapter 6 summarizes the overall argument, discusses certain issues that are outside the scope of the present research, and concludes.

Chapter 2

Locality Relations in Consonant Harmony

In this chapter, I first show that the typology of locality relations in consonant harmony (Section 2.1), or long-distance consonant assimilation, is associated with a cognitive learning bias, as evidenced by the results of an artificial language learning study of liquid harmony (Section 2.2). I then outline the Agreement by Correspondence (ABC) framework (Rose and Walker, 2004), arguing that satisfactory explanations of both the typology and the properties of human learning are tentatively offered by the implementation of ABC in Optimality Theory (Section 2.3). I note that the range of patterns considered in this chapter is intentionally restricted to those that are easily handled by a basic set of ABC constraints. This is done in order to exemplify what I consider to be an ideal scenario, in which an observed typological generalization that is demonstrably associated with a human learning bias is incorporated into the theory as a phonological constraint that seems to account for all of the empirical data. The reader is asked to bear in mind, however, that the unified account of locality relations that is presented in this chapter is not sustainable, and that Chapter 3 will highlight a number of problematic predictions that arise when the ABC approach to long-distance phonotactics is extended to more complex cases of consonant harmony and to consonant dissimilation.

2.1 Unbounded and Transvocalic Consonant Harmony

As outlined in Section 1.3, typological surveys of long-distance consonant agreement (Rose and Walker, 2004; Hansson, 2010a) reveal several interesting cross-linguistic properties. This chapter focuses primarily on a robust dichotomy that is observed with respect to the locality of the dependency, or the maximum distance between two interacting consonants. First, there are *unbounded* dependencies that hold in all …C…C… contexts, across any number of intervening segments of any kind. The second type of consonant harmony holds only within a relatively local …CvC… window. Following Hansson (2010a), throughout this dissertation I will refer to this type of locality as *transvocalic*, which is defined in terms of a dependency that hold across maximally one vowel (short or long) and never across an intervening consonant. This characterization stands in contrast to Rose and Walker (2004) and Bennett (2013), who follow Odden (1994) in referring instead to syllable-adjacency as the relevant context (see Section 1.3).

The data in (1) and (2) provide examples of sibilant harmony from Aari and Koyra, two related Omotic languages of Ethiopia that exhibit the difference between unbounded and transvocalic locality.

In Aari (see Hayward, 1990), the perfective suffix /-s/ surfaces faithfully as [-s] when no [–anterior] sibilant such as [ʃ] or [ʒ] precedes it, as seen in (1a). However, (1b)-(1f) demonstrates a pattern of sibilant harmony, in that the suffix surfaces instead as [-ʃ] when a [–anterior] sibilant precedes it. Note that there appears to be no upper limit on the trigger-target distance (Hansson, 2010a). For example, (1e) and (1f) show that the dependency holds across an intervening …VCVC… sequence.

(1) Unbounded sibilant harmony in Aari (Hayward, 1990)

a.	/baʔ-s-e/	ba**ʔs**e	‘he brought’
b.	/ʔuʃ-s-it/	ʔu**ʃʃ**it	‘I cooked’
c.	/tʃ’a̤ːq-s-it/	**tʃ**’a̤ːq**ʃ**it	‘I swore’
d.	/ʒaʔ-s-it/	**ʒ**aʔ**ʃ**it	‘I arrived’
e.	/ʃed-er-s-it/	**ʃ**eder**ʃ**it	‘I was seen’
f.	/ʒa̤ːg-er-s-e/	**ʒ**a̤ːger**ʃ**e	‘it was sewn’

In Koyra (see Hayward, 1982), the 3mSg (perfective) suffix /-osːo/ surfaces faithfully as [-osːo] when no [–anterior] sibilant precedes it, as seen in (2a). In (2b) and (2c), the suffix surfaces instead as [-oʃːo]—an alternation that is triggered whenever two sibilants are in a transvocalic context, separated by at most one vowel and no intervening consonants. The data in (2d) and (2e) provide evidence that the dependency holds only within the transvocalic window, as the alternation is not triggered when additional material intervenes.

(2) Transvocalic sibilant harmony in Koyra (Koorete; Hayward, 1982)

a.	/tim-d-osːo/	tindo**s**ːo	'he got wet'
b.	/patʃ-d-osːo/	pa**tʃ**ːo**ʃ**ːo	'it became less'
c.	/giːʒ-d-osːo/	giː**ʒ**ːo**ʃ**ːo	'it suppurated'
d.	/ʃod-d-osːo/	**ʃ**odːo**s**ːo	'he uprooted'
e.	/ʔatʃ-ut-d-osːo/	ʔa**tʃ**utːo**s**ːo	'he (polite) reaped'

Another example of the attested dichotomy is provided in (3) and (4), with data from two Bantu languages, Yaka and Lamba, spoken primarily in the Democratic Republic of the Congo and Zambia, respectively, that exhibit the difference between unbounded and transvocalic locality for a different type of long-distance interaction: nasal consonant harmony.[1]

As shown in (3), Yaka (see Hyman, 1995) has a perfective suffix /-idi/ that surfaces as [-ini] when it is preceded at any distance by a nasal consonant within the stem.

(3) Unbounded nasal consonant harmony in Yaka (Hyman, 1995)

a.	/-tsúb-idi/	-tsúbi**d**i	'wandered (perf.)'
b.	/-tsúm-idi/	-tsú**m**i**n**i	'sewed (perf.)'
c.	/-mák-idi/	-**m**áki**n**i	'climbed (perf.)'
d.	/-míːtuk-idi/	-**m**íːtuki**n**i	'sulked (perf.)'

A similar alternation (perfective /-ile/ → [-ine]) is demonstrated by the data in (4) for Lamba (data from Odden, 1994). However, while the alternation is triggered in

[1]Note that in patterns of Bantu nasal consonant harmony, intervening vowels do not surface with nasalization. That is, the production of (2b) is [-tsúmini], not *[-tsúmĩni]. For this reason, Rose and Walker (2004) and Hansson (2010a) present the patterns as cases of true long-distance agreement that is not achieved by spreading nasality (phonetically or phonologically).

transvocalic contexts, such as that of (4b), the dependency does not hold at greater distances, as seen in (4c).

(4) Transvocalic nasal consonant harmony in Lamba (Odden, 1994)

a. /-pat-ile/ -pati**l**e ‘scolded (perf.)’

b. /-uːm-ile/ -uː**m**i**n**e ‘dried (perf.)’

c. /-mas-ile/ -**m**asi**l**e ‘plastered (perf.)’

The above cases of transvocalic consonant harmony in (2) and (4) can be thought of as long-distance dependencies that are *bounded* in terms of distance; the pattern still holds when two elements of the dependency are non-adjacent, but only if they are in a relatively local …CvC… context. Beyond the transvocalic window, in …C…c…C… contexts, the restriction does not apply since the intervening material exceeds the transvocalic threshold for locality in that it contains one or more consonants. Of particular interest is the fact that there are no attested cases of long-distance consonant agreement that are bounded by any other measure (e.g., sibilant harmony across at most one consonant, two vowels, five segments, or any other metric of distance). Likewise, there are no examples of a language with consonant harmony that holds across exactly one intervening consonant (not more or less), or across at least one consonant (i.e. in *beyond-transvocalic* contexts; see Section 3.1.2.2 for more on beyond-transvocalic locality). These typological generalizations are summarized in Table 2.1.

Table 2.1: Attested and unattested variants of consonant harmony locality

Locality	Status	…CvC…	…CvcvC…	…CvcvcvC…
unbounded	attested	+	+	+
transvocalic	attested	+	–	–
$\leq$ 1 consonant	unattested	+	+	–
$=$ 1 consonant	unattested	–	+	–
$\geq$ 1 consonant	unattested	–	+	+

The remainder of this chapter investigates this dichotomy in more depth, providing evidence that it is associated with a human learning bias (Section 2.2), and summarizing how it has been incorporated into theory as a phonological constraint in the Agreement by Correspondence framework (Section 2.3).

Finally, note that in cases where both the unbounded and transvocalic versions of consonant harmony are represented within the same group of related languages, such as in the Omotic and Bantu cases mentioned above, it can often be concluded from independent evidence (e.g. geographic distributions) that the unbounded version of the sound pattern represents a secondary historical development from what was originally a strictly transvocalic dependency (Dolbey and Hansson, 1999; Gunnar Ólafur Hansson, pers. comm.). This entails a diachronic process of imperfect learning (overgeneralization) at some point in the history of the language(s) in question. While the exact nature of the issues surrounding a phonological change of this sort are not considered in depth in this dissertation, I point out that hints of overgeneralization from transvocalic to unbounded patterns are likewise evidenced in the results of the experiments presented throughout this dissertation.

2.2 Consonant Harmony in Artificial Language Learning

The typological locality universal described in the previous section has recently been reproduced in a laboratory setting by Finley (2011, 2012), who shows that, in the face of insufficient evidence, learners generalize in ways that adhere to the typology. In a set of artificial language learning experiments, subjects were tasked with learning sibilant harmony from a restricted set of training items, in which the choice of a suffix allomorph ([-su] or [-ʃu]) was dependent on a sibilant in the stem ([s] or [ʃ], respectively). To summarize, subjects exposed to "first-order" harmony in CVSV-SV (where S represents a sibilant) showed evidence of learning transvocalic harmony, but did not generalize to novel "second-order" SVCV-SV forms (Finley, 2011, Experiment1). By contrast, subjects learning harmony from …SVCV-SV contexts did generalize the dependency, both inwards to CVSV-SV contexts (Finley, 2011, Experiment 2) and outwards to SVCVCV-SV (Finley, 2012, Experiment 2).

McMullin and Hansson (2014, Experiment 1) produced a cohesive replication of Finley (2011, 2012) that incorporated several modifications to the experimental design. Specifically, the language had two suffixes, [-su] and [-ʃi], corresponding to "past" and "future" tense, respectively. These triggered a suffix-to-stem sibilant harmony pattern, such that any sibilant in the CVCVCV verb stem would alternate to match the suffix sibilant. The pattern thus differed from those in Finley (2011, 2012) in two respects: regressive directionality (from suffix to stem) and harmony alternations in open-class items (stems) rather than among a closed set of suffix allomorphs. The results aligned with those of Finley's experiments, extending and strengthening the overall evidence that the locality properties of attested long-distance phonotactics are shaped by a learning bias.

One motivation for using a sibilant contrast in studies of long-distance phonotactic learning is that, cross-linguistically, sibilant harmony is by far the most commonly attested type of non-adjacent consonant assimilation (Gafos, 1999; Hansson, 2010a). In this section, I present a replication of Finley (2011, 2012) and McMullin and Hansson (2014, Experiment 1), in which the target pattern is liquid harmony, a robustly attested (e.g. Bukusu; Odden, 1994, see data in Section 1.3) though less common type of consonant harmony. While this experiment serves to extend previous findings to a different and perhaps less acoustically salient featural contrast, it is also explicitly designed to allow for comparison along different dimensions, such as contrasting the learning of harmony vs. dissimilation patterns, through a series of additional experiments presented in subsequent chapters of this dissertation.

2.2.1 Experimental Methodology

This section describes the aspects of the experimental methodology that apply generally to all experiments presented in this dissertation. Details specific to Experiment 1, a study of liquid harmony, are presented below in Section 2.2.2.

2.2.1.1 Recruiting Participants

All participants were recruited through a subject pool made up of students at the University of British Columbia, and were eligible to sign up for just one of the experiments (i.e. there is no single person who is included in more than one training condition). Each participant either was compensated with $10 or received course credit for taking part in the experiment, which took approximately 45 minutes to complete. There were no restrictions on the language background of a participant, but the results presented below only include data from those who are self-reported native speakers of English. Note that many of the participants whose data were retained for analysis were not monolingual English speakers, but that no participant had experience with another language that is known to exhibit a long-distance consonant interaction.

2.2.1.2 Stimuli

The entire list of stimuli, which was designed to be used for Experiments 1 through 4, consisted of 1560 items. This included "verb" stems, as well as suffixed versions of each stem, where the suffixes [-li] and [-ɹu] correspond to "future" or "past" tense, respectively. The breakdown of these items is summarized in Table 2.2 (see Appendix A for the full list of stimuli), and Section 2.2.1.3 describes the stimuli in the context of each phase of the experiment. The stimuli were divided into four counterbalanced lists of 390, which were randomized and recorded in a soundproof booth by four phonetically trained native English speakers (2 male, 2 female). The speakers were not made aware of the fact that the list of stimuli adhered to any phonotactic pattern, and upon being asked to identify any patterns in the data, none of the speakers suggested the possibility of an interaction between the liquid consonants. Each speaker was instructed to produce each stimulus with word-initial stress, without vowel reduction, and all segments as they would in normal speech (e.g. mid vowels as diphthongs [eɪ] and [oʊ]). The stimulus set was designed such that each of the four speakers produced an equal number of each consonant and vowel occurring in each position of the three-syllable verb stems for each of three phases, as described below.

2.2.1.3 Experimental Design: Three Phases

All participants in Experiments 1 through 4 completed three phases of the experiment: a practice phase (identical for all groups), training (differed by group), and testing (identical for all groups).

Practice Phase A set of 8 verb stems (along with their two suffixed forms) was constructed for a practice phase, in which all participants (regardless of training condition) learned how to conjugate the verbs of the artificial language in past

Table 2.2: Breakdown of stimuli used in Experiments 1 through 4, where L denotes a liquid [l, ɹ], C is one of [p, t, k, b, d, g, m, n], and V is a vowel [i, e, o, u].

Phase	Stem type (#)	Suffixed forms (#)	Total # of stimuli
Practice	CVCV (8)	CVCV-li (8)	24
		CVCV-ɹu (8)	
Training	CVCVCV (96)	CVCVCV-li (96)	288
		CVCVCV-ɹu (96)	
	CVCVLV (96)	CVCVlV-li (96)	480
		CVCVɹV-li (96)	
		CVCVlV-ɹu (96)	
		CVCVɹV-ɹu (96)	
	CVLVCV (96)	CVlVCV-li (96)	480
		CVɹVCV-li (96)	
		CVlVCV-ɹu (96)	
		CVɹVCV-ɹu (96)	
Testing	CVCVLV (32)	CVCVLV-li (32)	96
		CVCVLV-ɹu (32)	
	CVLVCV (32)	CVLVCV-li (32)	96
		CVLVCV-ɹu (32)	
	LVCVCV (32)	LVCVCV-li (32)	96
		LVCVCV-ɹu (32)	
			Total = 1560

and future tense. During this phase, the participants first listened, over a set of headphones, to pairs of words consisting of a bare verb stem followed by its past tense form (e.g. [toke]…[toke-ɹu], [nipu]…[nipu-ɹu]). Note that, in contrast to the training phase described below, participants were not asked to repeat each of the words out loud in the practice phase. The interval time between the stem and its suffixed form was 500 ms. They then did the same for the future tense verbs (e.g. [toke]…[toke-li], [nipu]…[nipu-li]). To minimize any influence on the remainder of the experiment, verb stems in the practice phase were restricted to two-syllable CVCV stems with no liquids, where C represents a stop or nasal consonant [p, t, k, b, d, g, m, n] and V is one of four vowels [i, e, o, u]. The list of stimuli therefore included 24 individual items that were used in this phase (8 CVCV bare-stem forms, 8 CVCV-li future-tense forms, CVCV-ɹu past-tense forms).

Training Phase The training phase consisted of a series of 192 verb triplets. Each triplet was produced by the same speaker, and began with a three-syllable verb stem and was followed by its two suffixed forms with [-li] and [-ɹu]. This phase of the experiment was self-paced, with participants using the keyboard to advance through each of the items. Note that participants were asked to repeat each word aloud after hearing it, which is known to aid learning in similar tasks (e.g. Warker et al., 2009). The list of stimuli included 1248 words that were used in the training phase of the various training conditions. (Section 2.2.2.2 below describes the precise contents of the training phase for each of the three training conditions in Experiment 1. Similar sections are also included for subsequent experiments.)

One portion of this list included 96 CVCV$\underline{\text{L}}$V stems (to be used for the "Short-range" training conditions), where $\underline{\text{L}}$ represents a liquid consonant [l, ɹ]. Note that since the list of stimuli was designed to be used for all experiments in this dissertation (which look at patterns of both liquid harmony and liquid dissimilation), four suffixed forms were recorded for each of these stems. For example, a verb stem like [pide$\underline{\text{l}}$e] had two forms that adhered to a suffix-triggered pattern of liquid harmony (future tense [pide$\underline{\text{l}}$e-$\underline{\text{l}}$i] and past tense [pide$\underline{\text{ɹ}}$e-$\underline{\text{ɹ}}$u], and two that instead exhibited a dissimilatory pattern (future tense [pide$\underline{\text{ɹ}}$e-$\underline{\text{l}}$i] and past tense [pide$\underline{\text{l}}$e-$\underline{\text{ɹ}}$u]). For each Short-range training stem, the list of stimuli therefore included five total words (one stem and four suffixed forms), which resulted in 480 separate stimuli, which were counterbalanced for the number of stems with each liquid, the speaker who produced the stimuli for each set of stem+suffix items, and for the frequency of the non-liquid segments in each position of the stem.

Similarly, 96 "Medium-range" stems (CV$\underline{\text{L}}$VCV) were recorded along with each of their four possible suffixed forms. Each Medium-range stem was obtained by reversing the order of the second and third syllables of the Short-range stimuli (e.g. Short-range [kope$\underline{\text{ɹ}}$u] corresponded to Medium-range [ko$\underline{\text{ɹ}}$upe]). This maintained the counterbalancing for each of the same factors described above, and resulted in 480 different words that were also recorded for this portion of the set of stimuli.

The final contents of the training stimuli included 96 CVCVCV verb stems along with their past and future tense suffixed forms. Note that since these stems contained no liquids (e.g. [dutebi]), only two suffixed versions of each stem were required—one past tense ([dutebi-ɹ̲u]) and one future tense ([dutebi-l̲i]). This portion of the list therefore contained a total of 288 words.

Testing Phase Finally, each participant completed a testing phase, which used a Two-Alternative Forced Choice (2AFC) paradigm to determine whether participants preferred liquid agreement (harmony) or disagreement (disharmony) at three different levels of locality: Short-range (CVCVL̲V-L̲V), Medium-range (CVL̲VCV-L̲V), and Long-range (L̲VCVCV-L̲V). On each of 96 trials (32 for each of the testing distance), subjects heard a verb stem that contained one of the two liquids, and were asked to choose the correct option from two possible suffixed forms, each with the same suffix. All three items in each trial were produced by the same speaker (one of the same four encountered in the training phase). For example, the stem for one Long-range trial was [ɹomuge], with the options of [lomuge-li] or [ɹomuge-li]. The interval time between the verb stem and the first suffixed option was 500 ms, and the time between the first and second suffixed forms was 250 ms. Subjects were given a maximum of 3 seconds after the onset of the second option to respond before receiving an error message indicating that no response was recorded. The set of stimuli used in the testing phase (96 triplets = 288 total) was counterbalanced for several factors, including which liquid the verb stem contained, which suffix was used, which speaker produced the stimuli in each trial, whether the first of second 2AFC option showed liquid harmony, whether or not an alternation was required to achieve harmony, and the number of non-liquid consonants and vowels in each position of the test stems. Examples of testing trials at each distance are provided in Table 2.3.

Table 2.3: Example of testing trials for all experiments

Locality level	Test stem	2AFC options
Short-range	pidole	pidole-ɹu…pidoɹe-ɹu
(CVCV$\underline{\text{L}}$V-$\underline{\text{L}}$V)	duteɹe	dutele-li…duteɹe-li
Medium-range	tuluge	tuɹuge-li…tuluge-li
(CV$\underline{\text{L}}$VCV-$\underline{\text{L}}$V)	miɹete	milete-ɹu…miɹete-ɹu
Long-range	lugoni	lugoni-ɹu…ɹugoni-ɹu
($\underline{\text{L}}$VCVCV-$\underline{\text{L}}$V)	ɹomuge	lomuge-li…ɹomuge-li

2.2.1.4 Procedures

Upon providing consent to take part in the experiment, each participant was led into a small room that contained a computer, a set of headphones and a microphone. The experimenter gave the following oral instructions to each participant:

> In this experiment, you're going to use the headphones to hear words from a language. In the first part of the experiment, there is a short section where you will learn a little bit about how the language works. In the second part of the experiment, you'll hear a word, repeat it out loud, and then you can press any key on the keyboard to hear the next one, and so on. The microphone is going to record during that part of the experiment, but those recordings won't be used for anything outside of this study. Then there is a final section where you'll be tested on what you've learned, and for that part you'll switch to using the button box. You're going to hear two options for the right answer. If you think the first one is right, press '1', and if you think the second one is right, press '2'. We ask that you use one finger from each hand when selecting your answer.

Specific instructions were repeated with on-screen text prior to each phase of the experiment. After completing the study, subjects completed a language-background questionnaire, and were then debriefed and given the opportunity to ask any questions about the study.

2.2.2 Experiment 1: Liquid Harmony

2.2.2.1 Participants

Forty-eight self-reported native speakers of North American English (36 female, 12 male, mean age 23) took part in Experiment 1, with 16 subjects assigned to each of the three training conditions described below.

2.2.2.2 Training Conditions for Experiment 1

Subjects in Experiment 1 were assigned to one of three groups that differed with respect to the type of verb stems encountered in training. There were two experimental groups, M-Harm (for "Medium-range harmony") and S-Harm (for "Short-range harmony"), as well as a Control group. Recall that the training phase consisted of a series of 192 training triplets (a three-syllable verb stem followed by its two suffixed forms with [-li] and [-ɹu]).

For the experimental groups, a full half of the training stems contained no liquids and therefore provided no evidence of harmony in their suffixed forms (e.g. [dutebi ~ dutebi-ɹ̱u, dutebi-ḻi]). The remaining half of the training stems for the M-Harm group were of the form CVḺVCV, where Ḻ represents a liquid [l] or [ɹ] in the second syllable, while the S-Harm group was instead exposed to verbs stems of the shape CVCVḺV. Depending on their group, subjects were exposed to the liquid harmony either across an intervening consonant (e.g. piḻede ~ piḻede-ḻi, piɹ̱ede-ɹ̱u) for M-Harm subjects) or across one vowel with no intervening consonants (e.g. gutoɹ̱o ~ gutoḻo-ḻi, gutoɹ̱o-ɹ̱u] for S-Harm subjects). For each training item in which the stem contained a liquid, one of the two suffixes triggered a liquid alternation resulting in harmony, whereas the other suffixed form already obeyed harmony by morpheme concatenation alone. Training triplets were divided into two blocks, and randomized by subject within each block. Examples of training items for the two experimental groups are provided in Table 2.4, along with the number of each type of item.

Control subjects for this experiment completed the same amount of training, but were not exposed to any stems containing liquids. Instead, in each of the two training blocks, they were given the full set of the 96 triplets with CVCVCV stems. (They

Table 2.4: Examples of training items for M-Harm and S-Harm groups in Experiment 1

Group	Training triplet	Type and number of items
M-Harm	…dutebi…dutebi-ɹu…dutebi-li… …mekotu…mekotu-li…mekotu-ɹu…	96 stems with no liquid
	…pil̲ede…pil̲ede-l̲i…piɹ̲ede-ɹ̲u… …nel̲ogi…neɹ̲ogi-ɹ̲u…nel̲ogi-l̲i…	48 stems with [l]
	…koɹ̲upe…kol̲upe-l̲i…koɹ̲upe-ɹ̲u… …guɹ̲oto…guɹ̲oto-ɹ̲u…gul̲oto-l̲i…	48 stems with [ɹ]
S-Harm	…dutebi…dutebi-ɹu…dutebi-li… …mekotu…mekotu-li…mekotu-ɹu…	96 stems with no liquid
	…pidel̲e…pidel̲e-l̲i…pideɹ̲e-ɹ̲u… …negil̲o…negiɹ̲o-ɹ̲u…negil̲o-l̲i…	48 stems with [l]
	…kopeɹ̲u…kopel̲u-l̲i…kopeɹ̲u-ɹ̲u… …gutoɹ̲o…gutoɹ̲o-ɹ̲u…gutol̲o-l̲i…	48 stems with [ɹ]

thus differed from the experimental groups in that they were exposed to the same triplets twice each.) Participants in the Control group therefore saw no evidence for or against liquid harmony in their training. The results of the Control group are expected to reveal any underlying biases that may influence subject performance during the testing phase (e.g. from some aspect of the experimental design or from any gradient phonotactic patterns exhibited by the English lexicon), and will serve as a baseline in the statistical analysis.

2.2.2.3 Results and Analysis

Results of Experiment 1 were analyzed with a mixed-effects logistic regression model implemented in *R* (R Core Team, 2014) using the *glmer* function included in the *lme4* package (Bates et al., 2014). In what follows, I provide a detailed description of the statistical model, which is summarized in Table 2.5.

The categorical dependent variable in this model is whether the subject chose the test item exhibiting liquid harmony on a particular trial. The fixed effects por-

Table 2.5: Summary of the fixed effects portion of the mixed-effects logistic regression for Experiment 1 (N = 4518; log-likelihood = –2235.4)

Coefficient	Estimate	*SE*	$Pr(>\|z\|)$
Intercept	-1.13651	0.26127	< 0.0001
Harmony Faithful	2.47228	0.28800	< 0.0001
Harmony Second	-0.50501	0.12784	< 0.0001
Medium-range	0.09456	0.16467	0.5658
Long-range	-0.18878	0.16318	0.2473
S-Harm	1.28500	0.29777	< 0.0001
M-Harm	1.10671	0.30310	0.0003
Medium-range × S-Harm	-1.02310	0.22904	< 0.0001
Long-range × S-Harm	-0.89644	0.22821	< 0.0001
Medium-range × M-Harm	0.22645	0.23169	0.3284
Long-range × M-Harm	-0.29497	0.22715	0.1941

tion of the model included, as main effects, the between-subjects variable of training group (M-Harm and S-Harm compared to the baseline Control group) and the within-subjects variable of trigger-target distance in the test item (Medium- and Long-range compared to the baseline Short-range items), and an interaction between Group and Distance. The model also includes two nuisance variables that contributed significantly to the model fit, labelled *Harmony Faithful* (whether the liquid in the option with harmony was faithful to the stem liquid, or whether it required an alternation to achieve harmony) and *Harmony Second* (whether the option with harmony was presented as the second member of the pair of 2AFC items). The random component consisted of by-subject intercepts and slopes for the same two nuisance variables, which are intended to offset individual tendencies for choosing harmony vs. disharmony, faithfulness vs. alternations, and the first vs. second 2AFC alternative.

The baseline reference (Intercept term) of this model can be thought of as the log odds of choosing harmony for a subject in the Control group responding to a Short-range trial in which the first 2AFC option involved harmony by means of an unfaithful alternation (e.g. pidole … pidoɹe-ɹu, pidole-ɹu). The negative Intercept term thus indicates that on a trial like this, a Control subject is much less likely than chance to choose harmony. This is likely due in large part to the fact that an unfaithful alternation discourages a choice of harmony, as evidenced by the relatively large positive estimate for Harmony Faithful, which indicates that the log-odds of choosing harmony when the liquid remains faithful to the stem (e.g. tuluge … tuɹuge-li, tuluge-li) are highly increased compared to when an alternation is required. Similarly, the estimate for Harmony Second indicates that subjects are slightly less likely to choose harmony when it is presented as the second of the two options. The estimates for the main effect of the Distance parameters show that the Control group is slightly more likely to choose harmony at Medium-range (as compared to the Short-range distance), and slightly less likely to do so at Long-range, though neither effect reaches significance. The main effects of S-Harm and M-Harm training show that both experimental groups are significantly more likely than the Control group to choose harmony at the Short-range distance. The significant interactions between both Medium-range and Long-range distances with the S-Harm group show (with negative estimates) that the S-Harm group does not seem to apply harmony outside of the Short-range window. By contrast, neither of the interactions of Medium- and Long-range distances with the M-Harm group approach significance, indicating that the M-Harm group treats all three distances equally.

Note that in the logit mixed model presented in Table 2.5, the Short-range distance is used as the baseline reference, meaning that the coefficients for Group $\times$ Distance interactions merely show whether the experimental groups enforced harmony to the same degree at the other distances as they did in Short-range contexts; this is not equivalent to the hypothesis under consideration. Instead, Table 2.6 contrasts each experimental group with the control group by comparing the odds of selecting the form with harmony ([l…l] or [ɹ…ɹ]) at each of the three distance levels tested.

Each number in the table represents an odds ratio; for example, the odds that a subject in the S-Harm group would choose the harmony-obeying form for items of the Short-range type (cvcv<u>L</u>v-<u>L</u>v) are more than three times (3.61) those of a subject in the Control group doing the same. The odds ratios in Table 2.6 were ex-

Table 2.6: Odds ratios comparing experimental groups to control group for choosing harmony with each of the three testing distances as model baselines. Contexts encountered in training are in boldface and all cells that reach significance are shaded.

	Type of test item (trigger-target distance)		
	Short-range (cvcvLv-Lv)	Medium-range (cvLvcv-Lv)	Long-range (Lvcvcv-Lv)
M-Harm vs. Control	3.02 ($p < 0.001$)	**3.80 ($p < 0.001$)**	2.25 $p \approx 0.007$
S-Harm vs. Control	**3.61 ($p < 0.001$)**	1.30 $p \approx 0.374$	1.47 $p \approx 0.187$

tracted from the fitted logit mixed model by exponentiating the relevant coefficient estimates for the S-Harm and M-Harm terms in the model. Thus, for example, the figures in the Short-range column correspond to the coefficient estimates reported in Table 2.5, where Short-range was taken as the reference level of the Distance parameter ($e^{1.285} \approx 3.61$, $e^{1.107} \approx 3.02$). Odds ratios for the other two columns were obtained by re-fitting the exact same model but with Medium-range and Long-range, respectively, serving as the reference level for Distance; this allows for a direct comparison with the Control group at each distance level.

Finally, to facilitate a visual interpretation of the results of Experiment 1, several plots depicting mean aggregated proportions of harmony choices for each group at each level of locality, as well as the range proportions demonstrated by individual subjects, are provided in Figures 2.1, 2.2, and 2.3. These are discussed below, with effect sizes discussed with respect to the OR values reported in Table 2.6.

The results in Figure 2.1 demonstrate that the experimental groups did show evidence of learning the harmony pattern at the trigger-target distance levels on which they were trained (those highlighted in boldface in Table 2.6: Medium-range for the M-Harm group, Short-range for the S-Harm group). Each group was significantly more likely than the Control group to opt for a harmony response for test items of the relevant type, with odds ratios of 3.80 and 3.61, respectively.

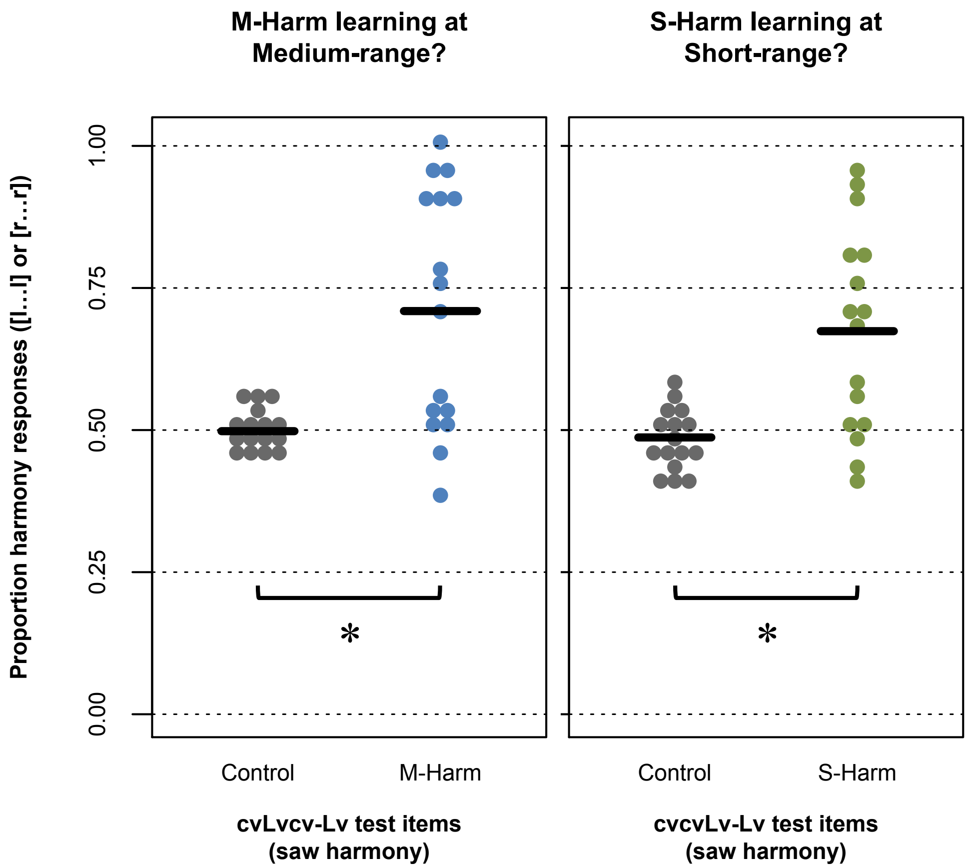

Figure 2.1: Plots comparing proportions of harmony responses for Control subjects to those of the M-Harm subjects in Medium-range test items (left panel) and to the S-Harm subjects in Short-range test items (right panel). Each dot represents individual subject performance, and group means are indicated with a horizontal line. Significance is extracted from a mixed logit model and indicates learning of the pattern each group was exposed to.

More importantly, as shown in Figure 2.2, subjects in the M-Harm group chose harmony more often than those in the Control group at the unfamiliar Short-range (OR=3.02) and Long-range (2.25) distances in the testing phase, and both effects reached statistical significance. This result, that the M-Harm group chose a test item with harmony significantly more often than the Control group for all levels of locality, is taken as evidence that subjects in the M-Harm training condition

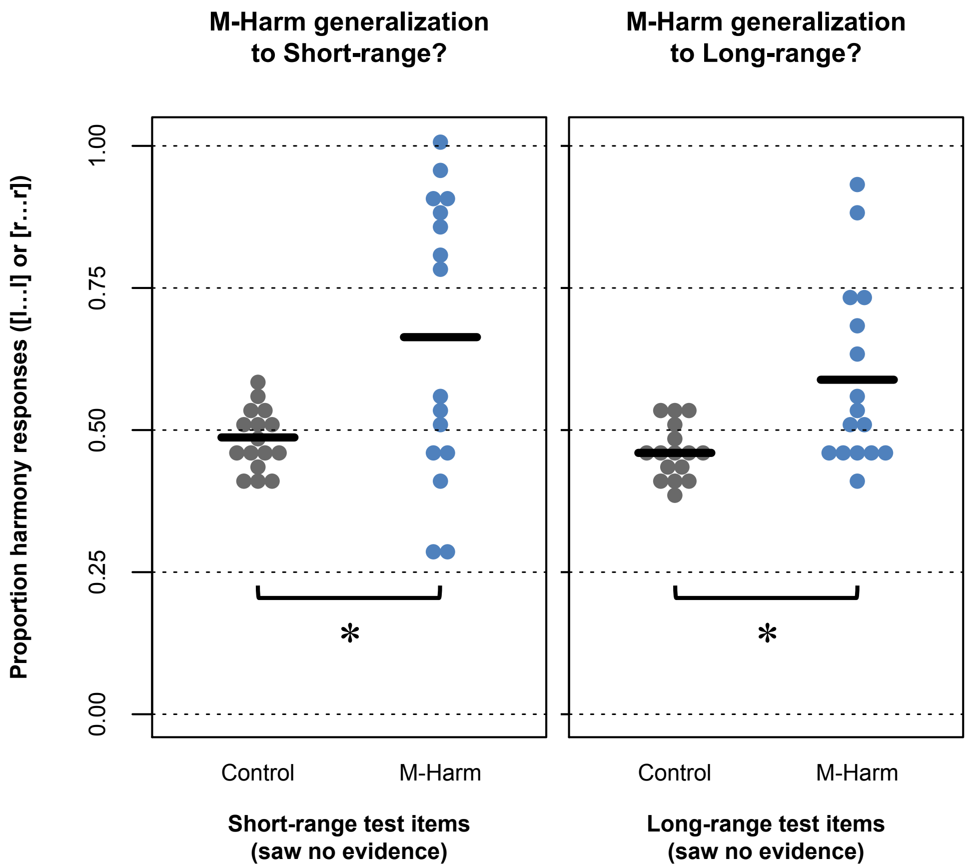

Figure 2.2: Plots comparing proportions of harmony responses for Control subjects to those of the M-Harm subjects in Short-range test items (left panel) and in Long-range test items (right panel). Each dot represents individual subject performance, and group means are indicated with a horizontal line. Significance is extracted from a mixed logit model and indicates generalization of the pattern the group was exposed to.

tend to interpret the dependency as unbounded liquid harmony, generalizing from a Medium-range pattern (which they were trained on and for which they showed evidence of learning) both inwards to Short-range and outwards to Long-range.

A similar effect was not seen for the S-Harm group, as shown in Figure 2.3. When taken as a whole, the S-Harm group does not seem to generalize the pattern of liquid harmony from the Short-range distance (which they were trained on, and

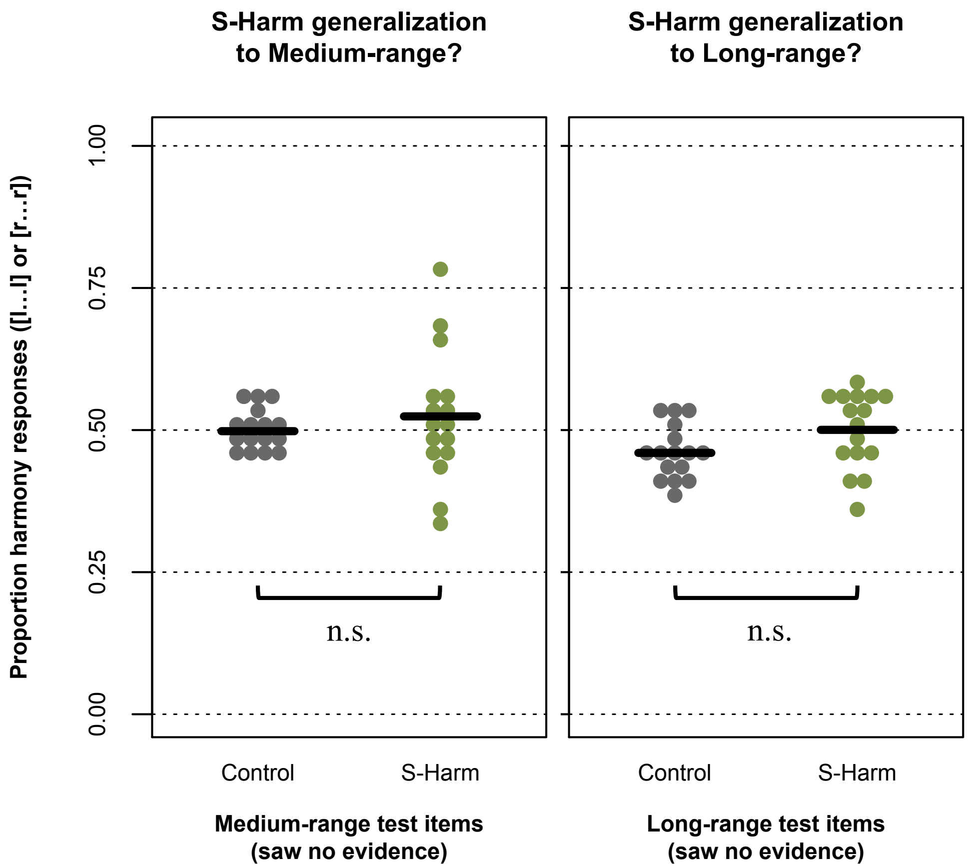

Figure 2.3: Plots comparing proportions of harmony responses for Control subjects to those of the S-Harm subjects in Medium-range test items (left panel) and in Long-range test items (right panel). Each dot represents individual subject performance, and group means are indicated with a horizontal line. (Non-)Significance is extracted from a mixed logit model and indicates generalization of the pattern the group was exposed to.

for which they showed evidence of learning) to Medium- (OR=1.30, $p \approx 0.374$) or Long-range (OR=1.47, $p \approx 0.187$) distances.

2.2.2.4 Summary and Discussion

As expected, the results of Experiment 1 replicate those of Finley (2011, 2012) and McMullin and Hansson (2014, Experiment 1), extending their findings to a different pair of segments (liquids [l, ɹ] rather than sibilants [s, ʃ]). To summarize, subjects who are exposed to consonant harmony in Short-range contexts (across just one vowel, with no intervening consonants) tend to interpret this evidence conservatively: they internalize a pattern that harmonizes transvocalic consonant pairs but does not extend to the more distant Medium- and Long-range contexts that were not encountered in training.[2] By contrast, subjects who are exposed to harmony in Medium-range contexts (spanning an intervening VCV sequence) tend to learn a dependency that holds at all locality levels, generalizing both inwards to Short-range and outwards to Long-range contexts. These results conform to the typological dichotomy among attested consonant harmony systems described in Section 2.1 above. I take this as evidence of a relationship between the limits of human phonotactic learning and the types of patterns observed in natural languages. Below I interpret these findings from the point of view of current phonological theory, showing that the treatment of locality in the Agreement by Correspondence framework straightforwardly captures both the typological generalizations and the properties of human learning with respect to patterns of consonant harmony.

2.3 Consonant Harmony as Agreement by Correspondence

While there exist many proposals for constraint-based analyses of consonant harmony, I focus here on the account provided by the Agreement by Correspondence framework (ABC; Walker, 2000a,c, 2001; Hansson, 2001, 2010a; Rose and Walker, 2004; Bennett, 2013), as this approach has seen relative success in accounting for the typology of consonant harmony. The ABC framework posits constraints that require pairs of segments to enter into a surface correspondence relation if they surpass some similarity threshold, due to sharing a certain set of feature values. For

[2] As can be seen in the above plots, there are a number of participants whose data do not reflect the overall trends of each of the experimental groups. Questions about individual differences, and the difference between "successful" learners vs. those who do not appear to learn anything are discussed at greater length in Section 3.2.3.

example, CORR[+strid] requires all co-occurring sibilants in the output to be surface correspondents of one another, regardless of how far apart they are in the word. The definition that I use for CORR[αF] (where [αF] is a set of feature specifications) is provided below in (5), which is a simplified version of the definition given by Bennett (2013, p. 55).[3]

(5) CORR[αF]: 'Two [αF] consonants must correspond.'
For each distinct pair of output consonants, X and Y, assign a violation if:
a. X and Y both have the feature specification [αF], and
b. X and Y are not in the same surface correspondence class

With the above definition, I follow Bennett (2013), who specifies that surface correspondence is an *equivalence relation*, in that it is symmetric (i.e. if X is a correspondent of Y, then Y is also a correspondent of X), transitive (i.e. if X is a correspondent of Y, and Y is a correspondent of Z, then X is a correspondent of Z), and reflexive (i.e. each member of a correspondence class is a correspondent of itself). While these properties are not of immediate consequence, Chapter 3 will demonstrate that transitivity in particular gives rise to a number of problematic predictions.

Other constraints (CC·Limiters; Bennett, 2013) impose restrictions on surface-corresponding segments. Among these are CC-IDENT[F] constraints, which require corresponding consonants to agree with respect to some feature (e.g. CC-IDENT[ant] requires agreement in [±anterior]) and thus serve as potential triggers and targets of harmony for that feature. Again, I follow Bennett (2013, p. 72), who defines CC-IDENT[F] as follows:

(6) CC-IDENT[F]: 'If two consonants correspond, then they agree on [±F].'
For each distinct pair of output consonants, X and Y, assign a violation if:
a. X and Y are in the same surface correspondence class, and
b. X is [αF], and
c. Y is [βF]
…where F is some feature, [αF], [βF] are its possible values, and α≠β

[3] The sole difference between the definition in (5) and Bennett's (2013) proposal is that Bennett specifies that the CORR constraint is assessed with respect to some morphological or phonological domain D, which for present purposes, I assume to be the word.

The final types of CC·Limiters that are relevant for the present discussion include ones that have been proposed as an account for the locality dichotomy of unbounded vs. transvocalic consonant harmony systems as previously discussed in Section 2.1. In particular, the constraint PROXIMITY (Rose and Walker, 2004) penalizes correspondence for any pair of consonants that are not within a bounded <u>C</u>V.<u>C</u> window. While Rose and Walker (2004, p. 494) define PROXIMITY in terms of syllable adjacency, I will use a modified version of their proposal that is framed with respect to transvocalic contexts:

(7) PROXIMITY: 'If two consonants correspond, then no consonant intervenes between them.'
For each distinct pair of output consonants, X and Y, assign a violation if:
a. X and Y are in the same surface correspondence class, and
b. there is some consonant Z that precedes Y and is preceded by X

Depending on the ranking of CORR, CC-IDENT, and PROXIMITY constraints (and additional CC·Limiters) relative to Faithfulness constraints (e.g. IO-IDENT[ant]), different variants of consonant harmony patterns can be generated.

2.3.1 Deriving Unbounded and Transvocalic Harmony in ABC

Tableau (8) shows a derivation of the Aari form in (1d), /ʃed-er-s-it/ → [ʃederʃit] 'I was seen', with an OT grammar for unbounded sibilant harmony. For expository reasons, in this and subsequent examples I abstract away from the issue of directionality of assimilation by assuming that any CC-IDENT constraint is specified for progressive or regressive directionality (sometimes indicated by C_LC_R-IDENT[F] and C_RC_L-IDENT[F], respectively; Rose and Walker, 2004). In any case, since the focus of this dissertation is on the proper characterization and learnability of the phonotactics themselves (permitted vs. prohibited output sequences), the question of which repair strategy emerges as optimal, and how this gets determined in the grammar, is not directly relevant.

(8) …S̲…c…S̲… sibilant harmony in Aari

/ʃed-er-s-it/	CORR [+strid]	CC-IDENT [ant]	IO-IDENT [ant]	PROXIMITY
a. $ʃ_x$e.der.$s_y$it	∗!			
b. $ʃ_x$e.der.$s_x$it		∗!		∗
c. ☞ $ʃ_x$e.der.$ʃ_x$it			∗	∗

Candidate (8a) loses as the two sibilants are not in correspondence, a violation of CORR[+strid]. Candidate (8b) is not optimal because its two sibilants, although in surface correspondence, have mismatched [±ant] specifications. The winner, then, is (8c), which incurs an IO-IDENT[ant] violation for the unfaithful mapping /s/ → [ʃ] in order to satisfy both the CORR and CC-IDENT constraints. With PROXIMITY ranked below CORR[+strid], the harmony will be enforced across any number of intervening segments or syllables. Note, however, that the constraint ranking in (8) would not be a suitable analysis of a transvocalic harmony pattern such as that exhibited by Koyra in (2), since in such cases harmony must be prevented from applying across an intervening consonant. Tableaux (9) and (10) give derivations of (2b) and (2c), demonstrating that a high-ranked PROXIMITY constraint results in harmony being enforced within the desired …C̲vC̲… window, but not beyond.

(9) …S̲vS̲… sibilant harmony in Koyra

/patʃ-d-osːo/	PROXIMITY	CORR [+strid]	CC-IDENT [ant]	IO-IDENT [ant]
a. patʃː$_x$osː$_y$o		∗!		
b. patʃː$_x$osː$_x$o			∗!	
c. ☞ patʃː$_x$oʃː$_x$o				∗

(10) No …S̲…c…S̲… sibilant harmony in Koyra

/ʃod-d-osːo/	PROXIMITY	CORR [+strid]	CC-IDENT [ant]	IO-IDENT [ant]
a. ☞ $ʃ_x$odːosː$_y$o		∗		
b. $ʃ_x$odːosː$_x$o	∗!		∗	
c. $ʃ_x$odːoʃː$_x$o	∗!			∗

The attested dichotomy between unbounded and transvocalic variants of consonant harmony locality can thus be accounted for with the relative ranking of Proximity and Corr[F] constraints. Assuming that a language has a pattern of consonant harmony in the first place, the dependency will be unbounded if Corr[F] ≫ Proximity, but will be transvocalic if Proximity ≫ Corr[F].

2.3.2 Biased Learning of ABC Constraint Rankings

This section explores the role that learning biases play in Optimality Theory and provides an account of the typological and experimental data as a reflection of ABC learning biases. The first type of bias to consider is one that shapes the boundary of the learner's hypothesis space. That is, we want to define a theoretical criterion for assessing whether a pattern is learnable or not. In OT terms, this is provided by the constraints themselves. The learner's goal is to discover the correct ranking of a finite number of constraints provided by a universal constraint set (CON). These constraints thus dictate the learner's hypothesis space, as patterns that cannot be generated by some ranking permutation (i.e. within the factorial typology) fall outside the region of learnable grammars. In Experiment 1, the pattern in question involves liquids agreeing for the [±lateral] feature, with the relevant constraints being Corr[liquid] (where [liquid] represents a feature bundle such as [+son, –nas, +cons] that picks out [l, ɹ] and omits all other segments), CC-Ident[lat], Proximity, and IO-Ident[lat]. With respect to locality when applying the pattern to two liquids, the twenty-four (four factorial) possible rankings of these constraints generate exactly the patterns observed in the typology: liquid agreement both in …LvL… and …L…c…L… contexts (unbounded harmony), agreement that is necessary only at the shorter …LvL… distance (transvocalic harmony), and free co-occurrence of liquids at any distance (no harmony). The crucial rankings for these patterns are presented in Table 2.7.

In summary, the set of constraints posited by the ABC account of consonant harmony can be thought of as learning biases, the factorial typology of constraints

[4] Though it is not relevant for present purposes, there are two types of faithful outputs that can emerge as optimal. Either the relevant pair of consonants will be in correspondence (but will not harmonize), resulting in a winning candidate such as […l$_x$a-ɹ$_x$u]) or they will not correspond at all (resulting in, for example, […l$_x$a-ɹ$_y$u]), though these two options are indistinguishable phonetically.

Table 2.7: Types of phonotactic patterns that can be generated within a factorial typology of ABC constraints.

Pattern	Crucial ranking	Number of rankings
Unbounded harmony	Corr[F] CC-Ident[G] Proximity IO-Ident[G]	5
Transvocalic harmony	Proximity Corr[F] CC-Ident[G] IO-Ident[G]	3
No restrictions (Faithfulness)[4]	any other ranking	16

corresponds to the types of grammars a learner must consider, and the prediction is that the set of possible phonotactics is exactly the set of patterns that can be generated by each of those grammars. There are many algorithms that can provably find a correct constraint ranking for a given pattern (e.g. Boersma, 1997; Tesar and Smolensky, 2000; Boersma and Hayes, 2001; Goldwater and Johnson, 2003), so long as it can be generated by at least one of the grammars in the factorial typology, and therefore both the unbounded and transvocalic variants of consonant harmony locality are learnable. Furthermore, other types of locality that are logically possible but unattested (such as those given above in Table 2.1) will not be learnable in this approach, since they cannot be generated using the set of constraints under consideration.

It is also important to know what the learner's first hypothesis is. In terms of OT, this is the initial constraint ranking, which can also be thought of as a type of learning bias. That is, in the absence of any evidence for changing the relative ranking of two or more constraints, which pattern would emerge as the output of the OT grammar? In order to ensure a restrictive final grammar, for example, there must be a bias that favours high-ranked Markedness constraints and low-ranked Faithfulness constraints (see, e.g., Smolensky, 1996; Hayes, 2004; Prince and Tesar, 2004; for similar arguments framed in terms of Harmonic Grammar, see Jesney and Tessier, 2011). Asking this question in the context of Experiment 1 will allow us to account for why the Control group did not apply harmony at any of the three testing distances, why the S-Harm group learned harmony in CVCV<u>L</u>V-<u>L</u>V contexts but did not generalize to other locality levels, and why the M-Harm group generalized from CV<u>L</u>VCV-<u>L</u>V contexts to all distances.

The fact that the Control group showed no preference for harmony at any of the three testing distances (and indeed largely made their choices based on which testing item was faithful to the bare verb stem) suggests that IO-IDENT[lat] begins as a highly ranked constraint with respect to the others.[5] This is likely a reflection of the presumed adult grammar of the native English speakers who participated in the experiment (see also Pater and Tessier, 2003). Based only on the results of the Control group, however, we have no means of determining the full ranking and must also look at the results of the other two groups.

Note that to have any pattern of liquid harmony, irrespective of locality, the crucial ranking is CORR[liquid], CC-IDENT[lat] $\gg$ IO-IDENT[lat]. In order for the learner to transition to a grammar that generates liquid harmony, then, IO-IDENT[lat] must be reranked such that it is dominated by both the CORR and CC-IDENT constraints. Which hypothesis the learner considers first (i.e. transvocalic vs. unbounded harmony) will depend on the relative ranking of the other constraints. If PROXIMITY outranks CORR[liquid] in the initial state, then the learner will hypothesize a transvocalic pattern until it is given exposure to harmony at a greater distance. If instead CORR[liquid] outranks PROXIMITY initially, then the learner will hypothesize an unbounded dependency until it is given explicit counter-evidence in beyond-transvocalic contexts. In Experiment 1, the S-Harm group was exposed to harmony only in CVCV<u>L</u>V-<u>L</u>V items—a pattern that is compatible with both types of locality—and did not generalize the pattern to greater distances. This suggests that PROXIMITY $\gg$ CORR[liquid] initially, but that the ranking can change if the learner is exposed to liquid harmony at a beyond-transvocalic distance, such as the CV<u>L</u>VCV-<u>L</u>V items seen by the M-Harm group in Experiment 1, resulting in

[5] It turns out that IO-IDENT[lat] does not need to be undominated for Faithfulness to be optimal at all distances. As long as it is ranked above either CORR[liquid] or CC-IDENT[lat], then an unfaithful candidate will never be optimal since it will always be better to leave the two liquids out of correspondence or to allow corresponding liquids to disagree.

a pattern with unbounded locality.

2.4 Summary and Conclusions

This chapter outlined a robust dichotomy with respect to the distance between two interacting segments in systems of consonant harmony. I then provided evidence from an artificial language learning experiment that the properties of human phonotactic learning reflect this typological dichotomy (unbounded vs. transvocalic locality). In the experiment, which replicates the results of similar studies looking at sibilant harmony (Finley, 2011, 2012; McMullin and Hansson, 2014), subjects in the M-Harm training condition were exposed to a pattern of liquid harmony that spanned an intervening non-liquid consonant (in cvLvcv-Lv contexts) and that group, as a whole, was significantly more likely than the Control group to apply the pattern to liquids at all three testing distances: Short-range (cvcvLv-Lv), Medium-range (cvLvcv-Lv), and Long-range (Lvcvcv-Lv) distances. By contrast, for subjects in the S-Harm training condition, who were exposed to cvcvLv-Lv harmony, the group applied the pattern in the Short-range test items, but was no more likely than the Control group to apply harmony at Medium-range and Long-range distances. The overall result was interpreted as evidence for a cognitive learning bias that imposes a restriction on the types of patterns that are human-learnable.

Upon establishing this relationship between typology and human learning, I demonstrated how the constraints used in the Agreement by Correspondence framework can account for the empirical data. Specifically, the factorial typology of the ABC constraints presented in this section includes consonant harmony patterns with both the unbounded and transvocalic variants of locality (but no other type of consonant harmony). This provides a straightforward explanation of the result that subjects in the M-Harm training condition of Experiment 1 tended to interpret the training stimuli as evidence for a pattern with unbounded locality—no other pattern that is compatible with harmony in cvLvcv-Lv contexts can be generated with these constraints. Furthermore, the fact that the S-Harm group did not show evidence of learning an unbounded dependency can be construed as a relatively high prior ranking of the Proximity constraint with respect to Corr[liquid].

Recall, however, that the purpose of this chapter was to present an intentionally restricted view of long-distance consonant interactions. The result is an illustration of what I consider to be the ideal scenario, in which a clear relationship between the typology and learnability of a phonotactic pattern is easily captured within a particular theoretical framework. In what follows, I will demonstrate that the ABC framework actually makes a number of problematic predictions, and argue for a new approach that is framed in terms of formal language theory.

Chapter 3

Problems with the ABC Approach

The purpose of this chapter is to call into question the validity of a number of predictions made by the Agreement by Correspondence (ABC) model of long-distance phonology (Hansson, 2001, 2010a; Rose and Walker, 2004; Bennett, 2013). I first point out that the factorial typology generated by ABC constraints, even in its simplest form, includes several phonotactic patterns that are unattested in natural language. With respect to long-distance consonant agreement—the types of patterns for which the framework was originally intended—the seemingly simple constraints used to account for the dichotomy of unbounded vs. transvocalic locality generate many pathological patterns that are unattested cross-linguistically (Section 3.1.1). Section 3.1.2 then presents a further problem that arises when extending the ABC framework to analyses of long-distance consonant disagreement (Bennett, 2013). The fact that both consonant harmony and dissimilation have similar cross-linguistic properties with respect to locality suggests that we should want to provide a unified theoretical account for the two types of patterns. However, an attempt to account for both within ABC results in unavoidably different predictions about each—in situations where harmony is enforced, dissimilation is expected to occur in exactly the complement set of environments (see Section 3.1.2). This is Bennett's 2013 "Mismatch Prediction". Though Bennett does provide limited typological support for the predicted case of *beyond-transvocalic* dissimilation (a complex interaction of liquids in Sundanese infixation; see also Cohn, 1992; Bennett, 2015), the second portion of this chapter explores the characteristics of human learning and generalization, demonstrating that subjects in an artificial language learning study (analogous to Experiment 1; see Section 2.2.2) do not differ when the target pattern is liquid dissimilation rather than liquid harmony (Section 3.2.2). With support from an additional analysis of Experiments 1 and 2 that includes only those subjects who surpassed a threshold that was used to indicate successful learning of the training pattern, I conclude that neither the typological nor experimental evidence supports the predictions of the ABC model of long-distance phonotactics and that we should pursue a different approach in order to provide a cohesive theoretical account of the observed empirical data.

3.1 Limited Typological Support for Predictions of ABC

3.1.1 Pathological Harmony Patterns

This section describes two pathological cases of consonant harmony. The first is what Hansson (2014) refers to as *agreement by proxy*, in which two relatively dissimilar consonants are forced to agree due to a shared relationship with a third (potentially non-harmonizing) correspondent (Section 3.1.1.1). The second problematic prediction is a sensitivity to the *count* (even vs. odd parity) of potential correspondents (Section 3.1.1.2), which prompts a new definition of the locality-based CC·Limiter PROXIMITY, which I call CC-CVC (modelled on CC·SYLLADJ, as defined by Bennett, 2013).

3.1.1.1 Agreement by Proxy

The assumption that surface correspondence is an equivalence relation (i.e. it is symmetric, reflexive, and transitive; Bennett, 2013), partitioning the set of co-occurring segments within the output into equivalence classes, can give rise to bizarre *agreement by proxy* effects (Hansson, 2014). Two co-occurring segments that are normally not required to correspond (nor, therefore, to interact) can be forced to do so when a third segment is present somewhere in the word, provided that this third segment is sufficiently similar to each of the other two to force them into (covert) correspondence with itself, and thereby also with each other. For example, we can imagine a hypothetical obstruent voicing harmony, in which only homorganic obstruent pairs interact, and where the harmony is moreover strictly regressive, harmonizing [–voice]…[+voice] sequences to [+voice]…[+voice] (while leaving [+voice]…[–voice] sequences intact). An input like /sada/ is thus changed to [zada], whereas a form like /saga/ surfaces faithfully as [saga]. A pattern like this is straightforwardly captured with the ranking in (1) and (2), where CORR[–son, αPlace] requires that co-occurring homorganic obstruents stand in correspondence and $\mathrm{C_RC_L}$-IDENT[+voi] penalizes [–voi]…[+voi] sequences of surface correspondents. The more general constraint CORR[–son], which demands correspondence for heterorganic as well as homorganic obstruent pairs, is ranked too low to have any effect.

(1) Obstruent voicing parasitic on place: homorganic pairs

/sada/	CORR [-son, αPlace]	$\mathrm{C_RC_L}$-IDENT[+voi]	IO-IDENT[voi]	CORR [-son]
a. $\mathrm{s_xad_ya}$	∗!			∗
b. $\mathrm{s_xad_xa}$		∗!		
c. ☞ $\mathrm{z_xad_xa}$			∗	

(2) Obstruent voicing parasitic on place: heterorganic pairs

/saga/	CORR [-son, αPlace]	$\mathrm{C_RC_L}$-IDENT[+voi]	IO-IDENT[voi]	CORR [-son]
a. ☞ $\mathrm{s_xag_ya}$				∗
b. $\mathrm{s_xag_xa}$		∗!		
c. $\mathrm{z_xag_xa}$			∗!	

Let us now imagine that the same language has another highly ranked CORR constraint, which demands that segments that agree in both manner ([±continuant]) and voicing must also stand in surface correspondence to one another: CORR[αcont, βvoi]. In cases like (1) and (2), such a constraint has no bearing on the outcome, as all the relevant output candidates vacuously satisfy it. However, in a case like (3), where the same kind of /s…g/ sequence as in (2) co-occurs with a /x/ somewhere else in the word, we see how the mere presence of this /x/ causes regressive harmony to be triggered in the /s…g/ sequence.

(3) Transitive correspondence relation causes *agreement by proxy*

/sagaxa/	CORR [αcont βvoi]	CORR [–son αPlace]	C_RC_L-ID[+voi]	IO-ID [voi]	CORR [-son]
a. $s_xag_yax_za$	∗!	∗!			∗∗∗
b. $s_xag_yax_ya$	∗!				∗∗
c. $s_xag_yax_xa$		∗!			∗∗
d. $s_xag_xax_xa$			∗!		
e. ☞ $z_xag_xax_xa$				∗	
f. $z_xag_yax_ya$				∗	∗!∗

For a [s…g…x] sequence like in (3a)-(3d), one CORR constraint requires [g…x] to be in correspondence (both are velar obstruents) while the other requires the same of [s…x] (both are voiceless fricatives). The only way to satisfy both CORR constraints is to place all three segments into the same correspondence class (3c)-(3d), but this means that [s…g] are also in correspondence with each other, unlike in cases like (2). Consequently, the generalization is that the regressive voicing harmony applies to heterorganic obstruent pairs if and only if the word also happens to contain a third obstruent that agrees in place with one but in manner with the other. To my knowledge, nothing resembling this kind of pattern has ever been attested in a natural language. Intuitively, it seems unlikely that this is an accidental gap, but it nonetheless falls within the boundary of possible phonotactic patterns as defined by the factorial typology of ABC constraints. While I do not provide any experimental evidence that a pattern of agreement by proxy is not human-learnable, Section 4.3 argues on computational grounds that such dependencies should indeed fall outside of the learner's hypothesis space.

3.1.1.2 Sensitivity to Number and Parity of Potential Correspondents

Another problematic prediction is the result of the definition of PROXIMITY (Rose and Walker, 2004, p. 494; redefined in transvocalic terms both in Section 2.3, and in (4) below). For expository purposes, I will reproduce the way in which transvocalic (or syllable-adjacent) harmony is generated in ABC, for a hypothetical example of transvocalic sibilant harmony (analogous to the pattern found in Koyra; see Section 2.1) that holds within a bounded S̲vS̲ window, but not at greater distances. At first glance, this appears easily captured with a high-ranking PROXIMITY constraint, which is defined in (4) below (in transvocalic terms, rather than syllable-adjacency as in Rose and Walker, 2004).

(4) PROXIMITY: Correspondent consonants must not be separated by any intervening consonant.

In words where the two consonants are in a transvocalic context, note that CORR[+strid] can be satisfied without violating higher-ranked PROXIMITY; a correspondence relation is therefore established and agreement in [±ant] is enforced over that relation. This is illustrated below in (5) with a hypothetical input /paʃasa/.

(5) High-ranked PROXIMITY permits …S̲vS̲… harmony

/paʃasa/	PROXIMITY	CORR [+strid]	CC-IDENT [ant]	IO-IDENT [ant]
a. paʃ$_x$as$_y$a		∗!		
b. paʃ$_x$as$_x$a			∗!	
c. ☞ paʃ$_x$aʃ$_x$a				∗

By contrast, when additional material intervenes between the two sibilants, harmony is not enforced, as shown below in (6) for the input /ʃapasa/.

(6) High-ranked PROXIMITY prevents …S̲…c…S̲… harmony

/ʃapasa/	PROXIMITY	CORR [+strid]	CC-IDENT [ant]	IO-IDENT [ant]
a. ☞ ʃ$_x$apas$_y$a		∗		
b. ʃ$_x$apas$_x$a	∗!		∗	
c. ʃ$_x$apaʃ$_x$a	∗!			∗

In the tableau in (6), the two sibilants are located outside of a transvocalic window and correspondence is prohibited due to the violations of PROXIMITY for candidates (6b) and (6c). As a result, candidate (6a) [ʃ$_x$apas$_y$a] would emerge as optimal, since the two sibilants are not in correspondence and therefore the candidate does not violate top-ranked PROXIMITY. This is a seemingly simple solution for generating transvocalic consonant harmony, and is indeed how Rose and Walker (2004) analyze the transvocalic nasal consonant harmony of Ndonga (in contrast to its unbounded counterpart in Kongo).

However, since each collection of surface correspondents within an output form constitutes a set (an equivalence class; see Bennett, 2013), in which every member is a correspondent of every other member, things quickly become complex once the number of correspondents goes above two, even if each local pair of consonants is in adjacent syllables (straddling a single vowel and nothing else). We should expect harmony to apply in a stepping-stone fashion—this is indeed what happens in real cases of transvocalic consonant harmony, such as Koyra, as shown in (7).

(7) Stepwise <u>S</u>V<u>S</u> sibilant harmony in Koyra (Koorete; Hayward, 1982)

a.	/dʒaʃ/	dʒaʃ	'fear'
b.	/dʒaʃ-us-/	dʒaʃ-uʃ	'cause to fear'
c.	/dʒaʃ-us-esːe/	dʒaʃ-uʃ-eʃːe	'let him/them frighten (s.o.)!'

However, the tableau in (8) shows how things go wrong for cases involving three consonants of the relevant class. The same constraint ranking that enforces harmony in a /…ʃVsV…/ sequence like in (5) will fail to do so in a /…ʃVsVsV…/ sequence.

(8) Harmony fails with three sibilants

/paʃasasa/	PROXIMITY	CORR [+strid]	CC-IDENT [ant]	IO-IDENT [ant]
a. paʃ$_x$as$_y$as$_z$a		***!		
b. paʃ$_x$as$_x$as$_x$a	*!		**	
c. ☹ paʃ$_x$aʃ$_x$aʃ$_x$a	*!			**
d. paʃ$_x$aʃ$_x$as$_y$a		**		*!
e. ☞ paʃ$_x$as$_y$as$_y$a		**		

The problem in situations like (8) is that in a chain of two or more transvocalic pairs of consonants of the relevant type, placing all of the consonants in correspondence will always result in a violation of PROXIMITY, since at least two of the correspondents will necessarily be separated by one or more intervening consonants, such as the first and third sibilants in (8b)-(8c). Given the ranking PROXIMITY $\gg$ CORR (which as explained above is the defining property of transvocalic harmony), the optimal resolution is to leave one of the consonants out of correspondence. The choice of which consonant to leave out falls to lower-ranked considerations such as Faithfulness, as can be seen in (8d) vs. (8e). Similarly, an input like /paʃaʃasa/ will surface without harmony, because [paʃ$_x$aʃ$_x$as$_y$a], with the first two sibilants in correspondence, will do better on Faithfulness than [paʃ$_x$aʃ$_y$aʃ$_y$a] with correspondence (and hence harmony) between the second and third sibilants. The crucial factor is thus not the number of potential harmony targets but the overall number of sibilants in the sequence.

In fact, the nature of the pathology is even more bizarre, in that the key criterion is the *parity* of that number, where the predictions for the (non)application of harmony are different for odd-parity vs. even-parity cases. With PROXIMITY $\gg$ CORR, an even number of potentially-harmonizing consonants is best partitioned into a series of transvocalic correspondence pairs (…$C_x VC_x VC_y VC_y V$…). Harmony is therefore predicted to be enforced in such words only between the 1st and 2nd consonant in the sequence, as well as in the 3rd and 4th consonant, etc., whereas there should be no requirement for harmony between the 2nd and 3rd, or the 4th and 5th (etc.) consonants. For an odd number of consonants, by contrast, the optimal correspondence configuration is for one consonant to stand outside of correspondence with any of the others, and for the (even-parity) sets of consonants on either side of that consonant to be partitioned into individual, harmonizing correspondence pairs as described above. Just as in (8), the determination of which consonant is the "odd man out" will fall to lower-ranked considerations such as Faithfulness. Needless to say, no natural language displays anything remotely resembling such a sound pattern.

In part to avoid problematic predictions like this, Bennett (2013) explicitly redefines Rose and Walker's 2004 PROXIMITY constraint such that it is only violated when there is some non-correspondent consonant that intervenes. Bennett (2013) calls this constraint CC·SYLLADJ, which is defined as in (9).

(9) CC·SYLLADJ: ‘Cs in the same correspondence class must inhabit a contiguous span of syllables.’ (Bennett, 2013, p. 85)
For each distinct pair of output consonants X and Y, assign a violation if:
a. X and Y are in the same surface correspondence class,
b. X and Y are in distinct syllables, ΣX and ΣY
c. there is some syllable ΣZ that precedes ΣY, and is preceded by ΣX
d. ΣZ contains no members of the same surface correspondence class as X and Y

Before demonstrating that this alleviates the parity-sensitivity, a revised version of the above constraint, which I call CC-CVC, is provided in (10). The key departure from Bennett’s CC·SYLLADJ is that CC-CVC assigns violations based on transvocalic locality rather than syllable-adjacency.

(10) CC-CVC: For each distinct pair of output consonants X and Y, assign a violation if:
a. X and Y are in the same surface correspondence class,
b. there is some consonant Z that precedes Y, and is preceded by X
c. Z is not in the same surface correspondence class as X and Y

As demonstrated in the tableau in (11), which uses the same ranking as the above tableaux but replaces PROXIMITY with CC-CVC, the desired pattern can now be generated, since sequences of three or more correspondents no longer violate the locality-based CC·Limiter.

(11) Sequences of three correspondents harmonize

/paʃasasa/	CC-CVC	CORR [+strid]	CC-IDENT [ant]	IO-IDENT [ant]
a. $\text{pa}ʃ_x\text{as}_y\text{as}_z\text{a}$		∗!∗∗		
b. $\text{pa}ʃ_x\text{as}_x\text{as}_x\text{a}$			∗!∗	
c. ☞ $\text{pa}ʃ_x\text{a}ʃ_x\text{a}ʃ_x\text{a}$				∗∗
d. $\text{pa}ʃ_x\text{a}ʃ_x\text{as}_y\text{a}$		∗!∗		∗
e. $\text{pa}ʃ_x\text{as}_y\text{as}_y\text{a}$		∗!∗		

Note, however, that all segments in correspondence are meant to form an equivalence class, but that CC-CVC is evaluated only after separating the correspondents into a series of pairs with linear order (that is, not for every possible pair). I point out (with Hansson, 2014) that this fix to the problem begins to undermine the definition of the correspondence relation, making correspondence classes look more like tiers or projections of segments, which is precisely the type of representation that Chapter 4 will argue for.

3.1.2 Pathological Dissimilation Patterns

This section begins by outlining how the constraint machinery of the Agreement by Correspondence framework offers dissimilation as an alternative repair for violations of a phonotactic restriction (Section 3.1.2.1). The result, as pointed out by Bennett (2013), is that the ABC model makes a number of concrete predictions about the typology of dissimilation. I will focus primarily on two predictions about locality that I argue are not supported by empirical data. First, Section 3.1.2.2 demonstrates how basic ABC constraints can produce a pathological *beyond-transvocalic* pattern of dissimilation that applies only outside of the local $\underline{C}$v$\underline{C}$ window. Second, Section 3.1.2.3 shows that with these same constraints, it is impossible to reproduce simple patterns of transvocalic dissimilation that are widely attested cross-linguistically (and discusses amendments to the ABC theory that Bennett, 2013 proposes to address this problem).

3.1.2.1 Dissimilation in the ABC Framework

In the Agreement by Correspondence model, long-distance dissimilation emerges as a strategy to avoid satisfying an agreement requirement by making the two consonants less similar, thereby eliminating the need to have them in correspondence in the first place (Bennett, 2013). To illustrate this, consider a hypothetical language that requires homorganic consonants to be in surface correspondence (facilitated by CORR[αPlace]), and that demands agreement for [±constricted glottis] (e.g. plain vs. ejective stop) among corresponding consonants (CC-IDENT[c.g.]). With a low ranking for faithfulness to input specifications for [c.g.], the optimal repair strategy is consonant harmony, such that two homorganic consonants are either both plain or both ejective, so long as all other constraints remain unviolated. This is shown below in Tableau (12), in which candidate (12c) is the winner.

(12) Long-distance [c.g.] harmony

/t'amata/	CORR [αPlace]	CC-IDENT [c.g.]	IO-IDENT [c.g.]
a. $\text{t'}_{\text{x}}\text{amat}_{\text{y}}\text{a}$	∗!		
b. $\text{t'}_{\text{x}}\text{amat}_{\text{x}}\text{a}$		∗!	
c. ☞ $\text{t'}_{\text{x}}\text{amat'}_{\text{x}}\text{a}$			∗

However, Tableau (13) shows that after demoting an IO-IDENT[place] constraint (which was previously assumed to be undominated), the candidate that exhibits harmony in (13c) is no longer the winner, and instead it is better to dissimilate the place feature in order to remove any need for correspondence at all, as seen in candidate (13d) below.

(13) Long-distance [place] dissimilation

/mat'ata/	CORR [αPlace]	CC-IDENT [c.g.]	IO-IDENT [c.g.]	IO-IDENT [place]
a. $\text{mat'}_{\text{x}}\text{at}_{\text{y}}\text{a}$	∗!			
b. $\text{mat'}_{\text{x}}\text{at}_{\text{x}}\text{a}$		∗!		
c. $\text{mat'}_{\text{x}}\text{at'}_{\text{x}}\text{a}$			∗!	
d. ☞ $\text{mat'}_{\text{x}}\text{ak}_{\text{y}}\text{a}$				∗

More generally speaking, by becoming less similar, a pair of consonants evades the scope of a high-ranked CORR constraint that would otherwise require them to be correspondents. Such avoidance can be driven by one or more high-ranked CC·Limiter constraints (Bennett, 2013, see Section 2.3) on surface correspondence configurations, if these cannot be satisfied in any other way (due to high-ranked Faithfulness, for example). This idea, that dissimilation happens only in situations where correspondence is penalized, whereas harmony can only take place where correspondence is permitted (since correspondence is the vehicle for agreement), gives rise to Bennett's (2013) "Mismatch Prediction" regarding the typology of these two families of sound patterns. That is, contexts that favour harmony should, other things being equal, be ones that fail to trigger dissimilation, and vice versa. Given that one type of CC·Limiter in the ABC theory penalizes any pair of correspondents that are located too far apart, such as PROXIMITY, CC·SYLLADJ, or CC-CVC (which are all defined above), these constraints too should be able to trigger dissimilation. As a result, we expect a typological mismatch between the locality patterns that are possible under consonant harmony and dissimilation, respectively. I note that while each of the locality-based CC·Limiters essentially makes the same problematic predictions, the remainder of this section uses CC-CVC to illustrate them.

3.1.2.2 Pathological Prediction: Beyond-Transvocalic Dissimilation

A locality-based CC·Limiter will trigger dissimilation only when the distance separating the relevant consonants is *greater* than the distance specified by the constraint (i.e. greater than a CvC distance for CC-CVC). The ABC model thus predicts the possibility of dissimilation patterns that are strictly *beyond-transvocalic*, applying in exactly the complement set of environments to what is seen for transvocalic consonant harmony.

The tableaux in (14) and (15) illustrate this schematically. As above, the hypothetical case in question involves surface correspondence between homorganic stops (CORR[αPlace]) and a demand that surface correspondents agree for the [c.g.] feature (CC-IDENT[c.g.]). The latter requirement can in principle be satisfied either directly, by laryngeal harmony under correspondence (/t'…t/ → [t'…t'], violating IO-IDENT[c.g.]), or indirectly, by place dissimilation out of correspondence (/t'…t/ → [t'…k], violating IO-IDENT[place]). However, as (14) shows, CC-CVC will, when ranked highly enough, trigger dissimilation on its own accord in beyond-transvocalic homorganic consonant pairs, even if CC-IDENT[c.g.] is too low-ranked to play any active role. For transvocalic pairs of consonants as in (15), on the other hand, correspondence is permitted and dissimilation is therefore not triggered. The result is dissimilation in beyond-transvocalic environments only. (As the comparison between (15b) and (15c) shows, such a dissimilation pattern can coexist either with harmony or faithful non-interaction in transvocalic contexts, depending on the ranking of the relevant CC-IDENT and IO-IDENT constraints.)

(14) Beyond-transvocalic dissimilation: CC-cvc triggers dissimilation

/t’amata/	Corr [αPlace]	CC-cvc	IO-Id [Place]	IO-Id [c.g.]	CC-Id [c.g.]
a. t’$_x$amat$_y$a	∗!				
b. t’$_x$amat$_x$a		∗!			∗
c. t’$_x$amat’$_x$a		∗!		∗	
d. ☞ t’$_x$amak$_y$a			∗		

(15) Beyond-transvocalic dissimilation: no CvC dissimilation

/mat’ata/	Corr [αPlace]	CC-cvc	IO-Id [Place]	IO-Id [c.g.]	CC-Id [c.g.]
a. mat’$_x$at$_y$a	∗!				
b. ☞ mat’$_x$at$_x$a					∗
c. ☞ mat’$_x$at’$_x$a				∗	
d. mat’$_x$ak$_y$a			∗!		

Unfortunately for the ABC model, this predicted mismatch between consonant harmony and dissimilation is a very poor fit for the attested typology. The only case that exhibits anything resembling a strictly beyond-transvocalic pattern, Sundanese rhotic dissimilation (Cohn, 1992; Bennett, 2013, 2015), is replete with other complications which make it far less persuasive as a test case (infixing morphology, co-existence with lateral harmony, sensitivity to stem-initial vs. non-stem-initial position, root vs. affix affiliation and onset vs. coda status).

3.1.2.3 Ranking Paradox: Basic Transvocalic Dissimilation

In contrast to the pathological beyond-transvocalic dependencies, which are easily generated by ABC constraints in spite of limited typological support, sound patterns in which dissimilation is confined to transvocalic locality are amply attested cross-linguistically (Odden, 1994; Bennett, 2013). However, no permutation of the ABC constraint types discussed thus far is capable of generating a dissimilation that is confined to transvocalic contexts without also applying at longer distances. The reason for this is that a ranking paradox arises in cases like those presented in

Tableaux (16) and (17) below.[1]

(16) CvC dissimilation requires Corr[F], CC-Ident[G] ≫ IO-Ident[F]

/mat'ata/	CC-cvc	Corr [αPlace]	CC-Ident [c.g.]	IO-Ident [Place]
a. mat'$_x$at$_y$a		∗!		
b. mat'$_x$at$_x$a			∗!	
c. ☞ mat'$_x$ak$_y$a				∗

The example in (16) shows that in order to have any dissimilation whatsoever in a …CvC… context, both Corr[αPlace] and CC-Ident[c.g.] must outrank IO-Ident[Place]. Otherwise, the output would remain faithful to the input form, with the homorganic stops either being out of surface correspondence, as in (16a), or entering a surface correspondence relation without repairing the sequence that violates the CC-Ident constraint, as in (16b). This holds true no matter what the ranking of the locality-based CC·Limiter is (i.e. CC-cvc), since it will never be violated by the …CvC… consonant pair.

However, with the necessary ranking to achieve …CvC… dissimilation being Corr[place], CC-Ident[c.g.] ≫ IO-Ident[Place], consonants in …C…c…C… contexts will also undergo dissimilation, no matter the ranking of CC-cvc, as shown in (17).

(17) Ranking implies dissimilation at all distances

/t'amata/	CC-cvc	Corr [αPlace]	CC-Ident [c.g.]	IO-Ident [Place]
a. t'$_x$amat$_y$a		∗!		
b. t'$_x$amat$_x$a	∗!		∗!	
c. ☞ t'$_x$amak$_y$a				∗

This is due to the fact that CC·Limiters may only penalize corresponding consonants and so candidates like (17c), which dissimilates in order to avoid correspon-

[1] The example in Tableaux (16) and (17) continues with the dissimilation of homorganic stops, but for clarity omits the IO-Ident[c.g.] constraint and does not offer laryngeal harmony as a possible repair.

dence altogether, will always be optimal. That is, it is impossible to rank these ABC constraints in a way that enforces transvocalic dissimilation without implying that dissimilation will hold at greater distances as well.

To deal with this problem, Bennett (2013) is forced to augment the model with special domain-restricted versions of the CORR constraints (without disposing of the locality-based CC·Limiters), which call for correspondence only in transvocalic consonant pairs. For example, CORR-CVC[αPlace] would penalize (16a) but not (17a). Replacing CORR[αPlace] with this constraint in (16) and (17) would produce a transvocalic-only dissimilation pattern, with winners (16c) and (17a), respectively. Such CORR-CVC[αF] constraints have previously been advocated by Hansson (2001, 2010b), but as an alternative to PROXIMITY/CC-SYLLADJ rather than complementary to it. The inclusion of both constraint types in the model creates an undesirable duplication of effort as well as rampant ambiguity of analysis, as practically every case of transvocalic consonant harmony can be interpreted either as involving the ranking CC-CVC≫CORR[αF] (with CORR-CVC[αF] ranked too low to be relevant) or else the undominated status of CORR-CVC[αF] (with a low ranking of CC-CVC and CORR[αF]).

3.2 Limited Experimental Support for ABC Predictions

3.2.1 Summary of Predictions

Predictions for Experiment 1 (liquid harmony; see Section 2.2.2) were obtained by assuming a relationship between the typology of phonotactic patterns and the way humans learn those patterns. Since there are no attested cases of harmony exhibiting anything other than unbounded or transvocalic locality, the hypothesis (which the results supported) was that the subjects in the M-Harm training group, who were learning from cv<u>L</u>vcv-<u>L</u>v contexts, would generalize to all levels of locality. Furthermore, since cross-linguistic evidence suggests that the unbounded patterns emerge from more local transvocalic dependencies, there was reason to suspect that subjects in the S-Harm training group would not generalize beyond the distance encountered in their cvcv<u>L</u>v-<u>L</u>v training items. This was also supported by the results, and the inductive biases shown by the learners were easily framed within the ABC model of consonant harmony, by varying the relative ranking of PROXIMITY.

Based only on the typology, we would make the same predictions for dissimilation, since unbounded and transvocalic variants of long-distance consonant dissimilation are both attested and, with but one possible exception (Sundanese; see Cohn, 1992; Bennett, 2013, 2015), no other types of locality are found cross-linguistically. While it seems fortuitous that the ABC framework facilitates dissimilation as an alternative repair for prohibited pairs of (non-adjacent) consonants, using the same constraint set gives rise to a set of predictions that contradicts the previous logic. As discussed above in Section 3.1.2, the factorial typology of ABC constraints includes sound patterns in which dissimilation is enforced only beyond the CvC window. Furthermore, in order to produce simple cases of transvocalic dissimilation, Bennett (2013) posits that the ABC model includes not only PROXIMITY, but an additional, and otherwise redundant, locality-based CORR constraint (CORR-CVC[αF]). There are thus at least two competing, yet both motivated, hypotheses for an experiment looking at the learnability of long-distance dissimilation.

If there is indeed a strong connection between the learnability of phonotactic patterns and the way such patterns are distributed cross-linguistically, then we would expect subjects to learn and generalize the pattern of dissimilation in the same way as subjects did in Experiment 1 when learning harmony. Subjects in an M-Diss training group would generalize the learned pattern to all distances, both shorter and greater than the CVLVCV-LV contexts encountered in training. S-Diss subjects on the other hand would tend to restrict the pattern to apply only in the types of Short-range CVCVLV-LV items that they were exposed to in training and not at greater distances.

The ABC model of long-distance phonotactics generates a different set of predictions. For the M-Diss group, there are two potential outcomes that fit the theory. Since the CVLVCV-LV dissimilation encountered in training is compatible with both the unbounded and beyond-transvocalic variants of locality for dissimilation that can be generated using basic ABC constraints, we would expect to see subjects either generalize to all distances (unbounded locality) or to generalize only to the greater LVCVCV-LV contexts, without enforcing the pattern for Short-range CVCVLV-LV test items (beyond-transvocalic locality). By contrast, since a 'strictly transvocalic' variant of locality is rather difficult to generate using ABC constraints, subjects in the S-Diss group would be expected to have some tendency (at least more so than the S-Harm group) to interpret the observed pattern as an unbounded dependency. These two contradictory sets of predictions are the motivation for extending the artificial language learning study to patterns of liquid dissimilation in Experiment 2.

3.2.2 Experiment 2: Liquid Dissimilation

3.2.2.1 Methodology

Participants, Stimuli, and Procedure Participants for two additional training conditions were recruited and compensated as described in Section 2.2.1, resulting in 32 new participants for Experiment 2 (16 female, 8 male, mean age 23), with 16 assigned to each of the "M-Diss" and "S-Diss" groups.

The details of the methodology for Experiment 2 are nearly identical to those of Experiment 1 (see Section 2.2.1), and all stimuli required for both experiments were recorded during the same sessions (one session for each of the four speakers). All participants completed a practice phase, a training phase, and a testing phase, with the sole difference being the type of pattern subjects in the experimental groups were exposed to during training.

Training Conditions All training stems were identical to those used in Experiment 1. However, rather than seeing alternations that resulted in liquid harmony, subjects in the M-Diss group were exposed to triplets that exhibited a pattern of liquid *dissimilation* in cvLvcv-Lv contexts. Subjects in the S-Diss group were also exposed to liquid dissimilation, but only for cvcvLv-Lv items. The testing phase in Experiment 2 was identical to that of Experiment 1. A breakdown of the number and type of stimuli presented in the training phases for the M-Diss and S-Diss groups is provided in Table 3.1.

Table 3.1: Examples of training items for M-Diss and S-Diss groups in Experiment 2

Group	Training triplet	Type and number of items
M-Diss	…dutebi…dutebi-ɹu…dutebi-li…	96 stems with no liquid
	…mekotu…mekotu-li…mekotu-ɹu…	
	…pil̲ede…piɹ̲ede-l̲i…pil̲ede-ɹ̲u…	48 stems with [l]
	…nel̲ogi…nel̲ogi-ɹ̲u…neɹ̲ogi-l̲i…	
	…koɹ̲upe…koɹ̲upe-l̲i…kol̲upe-ɹ̲u…	48 stems with [ɹ]
	…guɹ̲oto…gul̲oto-ɹ̲u…guɹ̲oto-l̲i…	
S-Diss	…dutebi…dutebi-ɹu…dutebi-li…	96 stems with no liquid
	…mekotu…mekotu-li…mekotu-ɹu…	
	…pidel̲e…pideɹ̲e-l̲i…pidel̲e-ɹ̲u…	48 stems with [l]
	…negil̲o…negil̲o-ɹ̲u…negiɹ̲o-l̲i…	
	…kopeɹ̲u…kopeɹ̲u-l̲i…kopel̲u-ɹ̲u…	48 stems with [ɹ]
	…gutoɹ̲o…gutol̲o-ɹ̲u…gutoɹ̲o-l̲i…	

3.2.2.2 Results and Analysis

This section presents the results of Experiment 2, analyzing them in much the same way as was done above for Experiment 1 (see Section 2.2.2.3 for full descriptions). As the two experiments were designed and implemented at the same time, the analysis presented here uses the response data from the same 16 Control subjects used for the analysis of Experiment 1. The preliminary mixed-effects logistic regression analysis of the results includes data from 48 subjects (the first 16 in each of the three conditions who completed the study) and models the log-odds of choosing the test item that had two *different* liquids. The fixed effects portion of the model includes the between-subjects variable of training group (M-Diss and S-Diss compared to the baseline Control group) and the within-subjects variable of trigger-target distance in the test item (Medium- and Long-range compared to the baseline Short-range items), and an interaction between Group and Distance. The model also includes two nuisance variables that contributed significantly to the model fit, labelled *Dissimilation Faithful* (whether the stem liquid in the option with disagreement was faithful to the liquid in the bare-stem form, or whether it required a dissimilatory alternation) and *Dissimilation Second* (whether the option with disagreeing liquids was presented as the second member of the pair of 2AFC items). The random component consisted of by-subject intercepts and slopes for the same two nuisance variables, which are intended to offset individual tendencies for choosing harmony vs. disharmony, faithfulness vs. alternations, and the first vs. second 2AFC alternative. This model, which uses the Short-range test items as its baseline for the Distance variable, is summarized in Table 3.2.

Table 3.2: Summary of the fixed effects portion of the mixed-effects logistic regression for Experiment 2 (N = 4534; log-likelihood = –2128.3)

Coefficient	Estimate	*SE*	$Pr(>\|z\|)$
Intercept	−0.72645	0.24415	0.0029
Dissimilation Faithful	2.38678	0.29248	< 0.0001
Dissimilation Second	−0.71357	0.13958	< 0.0001
Medium-range	−0.09605	0.16489	0.5602
Long-range	0.18617	0.16339	0.2545
S-Diss	2.51628	0.30460	< 0.0001
M-Diss	1.47546	0.30615	< 0.0001
Medium-range × S-Diss	−2.03104	0.24981	< 0.0001
Long-range × S-Diss	−2.70534	0.25122	< 0.0001
Medium-range × M-Diss	−0.00588	0.23654	0.9802
Long-range × M-Diss	−0.98671	0.23281	< 0.0001

The model in Table 3.2 leads to many of the same conclusions that were drawn from the analogous mixed logit model for the first 16 subjects in each condition of Experiment 1 (see Table 2.5). The negative estimate for the intercept (which uses all of the baseline measures for each of the variables) can be interpreted as follows: a subject in the Control group is less likely than chance to choose a Short-range test item with disagreeing liquids ([l…ɹ] or [ɹ…l]) when it requires a dissimilatory alternation and is presented first item of the 2AFC pair. The likelihood of preferring disagreeing liquids increases when the stem liquid does not require an alternation (*Dissimilation Faithful*), but decreases when it is the second item of the 2AFC choices (*Dissimilation Second*). The relatively small effects and lack of significance for the *Medium-range* and *Long-range* predictor variables indicates that a Control subject is not any more or less likely to choose dissimilation at either of these distances compared to the Short-range baseline. Main effects for both experimental groups (S-Diss and M-Diss) are positive and significant, suggesting that subjects in each of these two groups were more likely to choose dissimilation in Short-range test items than Control subjects were. Based on the large and significant negative coefficient estimates of the first two interaction terms (*Medium-range* $\times$ *S-Diss* and *Long-range* $\times$ *S-Diss*), a subject in the S-Diss group is much less likely to choose dissimilation at the Medium- and Long-range distances (which were not encountered in training), suggesting that they did not generalize outwards from the Short-range distance. Estimates for the final two interaction terms in the model (*Medium-range* $\times$ *M-Diss* and *Long-range* $\times$ *M-Diss*) indicate that the M-Diss subjects are statistically no less likely to choose dissimilation at Medium-range compared to the baseline Short-range, but that they are significantly less likely to do so at Long-range.

Recall from the discussion in Section 2.2.2.3 that a model of subject behaviour that uses a particular baseline measure of Distance cannot, strictly speaking, allow us to assess the hypotheses under consideration. As a more appropriate alternative, Table 3.3 provides a direct comparison, in the form of odds ratios, for each of the two experimental groups after re-fitting the same mixed-effects logistic regression model with different choices of the baseline level for the test-item Distance factor.

This table, along with Figure 3.1 illustrates that both experimental groups show evidence of learning. Overall, the M-Diss subjects learned a pattern of dissimilation from their cv$\underline{\text{L}}$vcv-$\underline{\text{L}}$v training items, as they were more than four times (4.35) more likely than the Control group to choose dissimilation in the novel Medium-range test items. Likewise, subjects in the S-Diss group appear to have learned the pattern they were exposed to, being more than twelve times (12.38) more likely than the Control group to choose dissimilation in the cvcv$\underline{\text{L}}$v-$\underline{\text{L}}$v context that they were exposed to in training.

As seen in Figure 3.2, the M-Harm group seems to generalize this pattern inwards to the Short-range test items (4.37 times more likely to choose dissimilation

Table 3.3: Odds ratios comparing experimental groups to Control group for choosing dissimilation with each of the three testing distances as model baselines. Contexts encountered in training are in boldface and all cells that reach significance are shaded.

	Type of test item (trigger-target distance)		
	Short-range (cvcvLv-Lv)	Medium-range (cvLvcv-Lv)	Long-range (Lvcvcv-Lv)
Nontransvocalic vs. Control	4.37 ($p < 0.001$)	**4.35 ($p < 0.001$)**	1.63 $p \approx 0.104$
Transvocalic vs. Control	**12.38 ($p < 0.001$)**	1.62 $p \approx 0.087$	0.83 $p \approx 0.498$

than the Control group), but the same effect was not seen for the Long-range test items. The M-Diss group showed a small increase (OR=1.63) in the likelihood of choosing disharmony, compared to the Control group, in test items with a stem-initial liquid (LVCVCV-LV items), but the effect did not reach statistical significance.

Finally, subjects in the S-Diss group are not significantly more likely to choose dissimilation at either of the greater distances, being about 1.62 more likely to choose dissimilation at Medium-range and about equally as likely (OR=0.83) as the Control group to choose dissimilation at Long-range. Neither of these effects reached statistical significance.

3.2.2.3 Discussion of Experiment 2

The goal of Experiment 2 was to create a study of long-distance phonotactic learning identical to Experiment 1 (in terms of the overall methodology) in order to evaluate two competing, yet both motivated hypotheses. It turns out, however, that the results of Experiment 2 (as presented in Section 3.2.2.2) are not compatible, at least in their entirety, with any of the predictions outlined in Section 3.2.1. As can be seen in Table 3.3, which presents the size and significance of the odds ratios that emerge from the statistical analysis that compared each group to the Control group

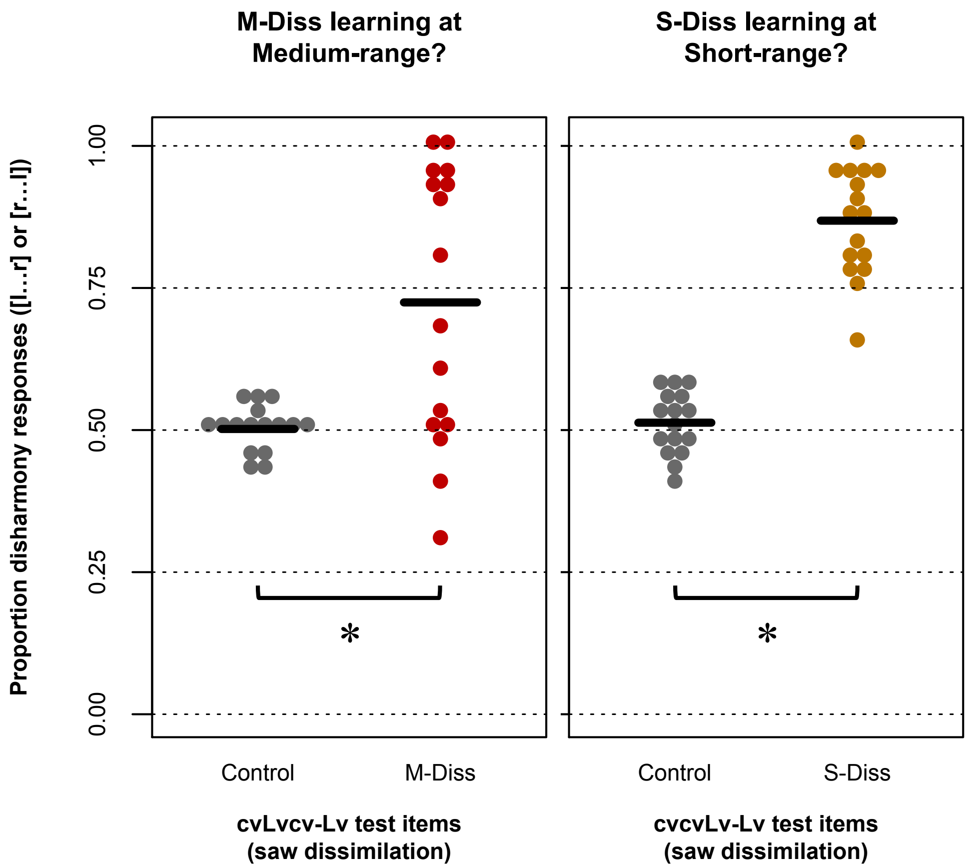

Figure 3.1: Plots comparing proportions of disharmony responses for Control subjects to those of the M-Diss subjects in Medium-range test items (left panel) and to the S-Diss subjects in Short-range test items (right panel). Each dot represents individual subject performance, and group means are indicated with a horizontal line. Significance is extracted from a mixed logit model and indicates learning of the pattern each group was exposed to.

at each of the three testing distances, subjects in the M-Diss group were significantly more likely to choose the items with disagreeing liquids at Short-range and Medium-range, but this result did not extend to Long-range test items. This is actually problematic for all of the considered hypotheses—M-Diss subjects were expected to either generalize the pattern from Medium-range to all distances, or only outwards to the Long-range test items in accordance with the beyond-transvocalic

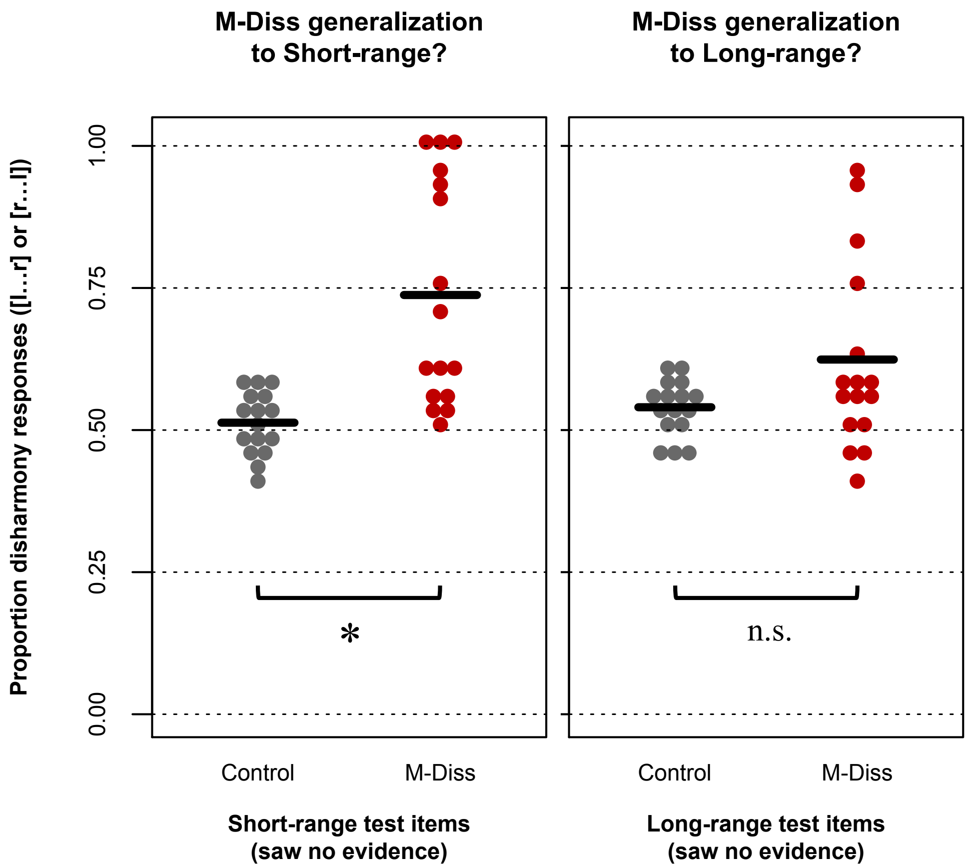

Figure 3.2: Plots comparing proportions of disharmony responses for Control subjects to those of the M-Diss subjects in Short-range test items (left panel) and in Long-range test items (right panel). Each dot represents individual subject performance, and group means are indicated with a horizontal line. (Non-)Significance is extracted from a mixed logit model and indicates generalization of the pattern the group was exposed to.

patterns predicted by ABC. Note, however, that the present results may be confounded by a more general reluctance for experimental subjects to extend phonological alternations into salient word-initial contexts[2] (see, e.g., Becker et al., 2012).

[2]The same reluctance was, to some degree, also seen in Experiment 1. The M-Harm group was less likely to choose the test item with harmony in Long-range test trials (OR=2.25; see Table 2.6) than for Short-range (OR=3.02) or Medium-range (3.80) trials. However, the resistance to initial-syllable alternations was not enough to reduce the effect below significance for the M-Harm group.

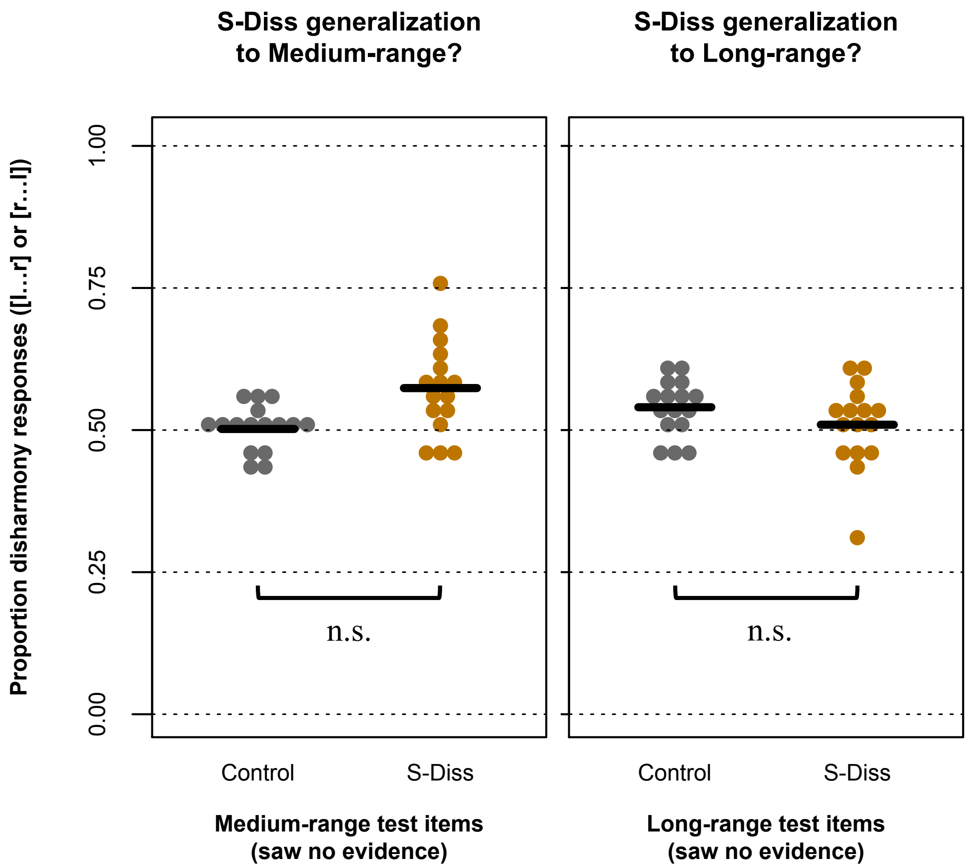

Figure 3.3: Plots comparing proportions of disharmony responses for Control subjects to those of the S-Harm subjects in Medium-range test items (left panel) and in Long-range test items (right panel). Each dot represents individual subject performance, and group means are indicated with a horizontal line. (Non-)Significance is extracted from a mixed logit model and indicates generalization of the pattern the group was exposed to.

With respect to the S-Diss group, subjects applied the pattern of dissimilation to the Short-range test items (indicating learning), but were not significantly more likely than the Control group to choose dissimilation at either of the Medium- or Long-range test items. This result is in line with the cross-linguistic distribution of non-adjacent consonant dissimilation, in that strictly-transvocalic locality is relatively common and does not necessarily imply dissimilation at greater distances. The lack of generalization to either of the greater distances goes against the predictions of the ABC model, however, as the basic constraint set does not permit dissimilation to apply at a transvocalic distance without implying unbounded locality.

3.2.3 Analysis of Successful Learners in Experiments 1 and 2

3.2.3.1 Motivation for Additional Analysis

Evidence from the analysis of both Experiments 1 and 2, as described above in Section 2.2.2.3 and Section 3.2.2.2, respectively, suggests that the results are in line with the hypothesis that human language learners only have access to two possible variants of locality for long-distance phonotactics: transvocalic or unbounded. These are the only two types of locality that are reliably attested cross-linguistically, but are not the only patterns the learner should have access to if the ABC framework is an accurate account of long-distance consonant agreement and disagreement. However, the evidence is not entirely definitive, especially with respect to Experiment 2, and there exist reasons to be skeptical of any conclusions.

The analyses previously presented for Experiments 1 and 2 were mixed-effects logistic regression models whose random effects structures were included in order to factor out individual subject tendencies in the test phase. However, when examining the plots of individual subject responses in Figure 2.1 and Figure 3.1 (which show the results for each experimental group at the testing distance that corresponds to the type of pattern encountered in training), it is clear that the proportion of (dis)harmony responses are not clustered around the mean. Instead, subjects tend to fall into one of two categories: 'successful learners' and 'non-learners'. Successful learners are those subjects whose responses place them closer to the 100% mark for the testing distance that corresponds to their training (i.e. Short-range for S-Harm and S-Diss subjects, Medium-range for M-Harm and M-Diss) and non-learners are those subjects who remain within or close to the range of the Control subjects (around 50%, usually due to choosing the 2AFC option that was faithful to the stem). While it is also an interesting question to ask "Which of the two types of locality is *more difficult* to learn" (perhaps measured by the proportion or number of subjects that successfully learned the target pattern), the goal of the present study is to determine whether or not human learners generalize patterns learned from an impoverished input in a way that matches the typology (or the theoretical predictions). With this in mind, the remainder of Section 3.2.3 presents an alternative and arguably more appropriate analysis of Experiments 1 and 2, augmented with data from further subjects, as described in the next section.

3.2.3.2 Defining a Threshold for Successful Learning

The preliminary analyses of Experiments 1 and 2 included the response data from the first 16 participants in each group. The following procedure was used to instead obtain 12 "successful" learners in each of the S-Harm, M-Harm, S-Diss, and M-Diss groups. A subject was considered to have reached the threshold for successful learning provided that the proportion of responses that adhered to the target pattern *at the same distance that the subject was exposed to in training* surpassed the 95% confidence level using a one-tailed test on a binomial distribution.[3] I note that this threshold was defined independently of the previously collected data (i.e. without knowing how many of the first 16 subjects in each condition would qualify as learners). Intuitively, the idea was to ensure that the probability of a subject who responded randomly being classified as a learner was less than 1 in 20. For example, if an M-Harm subject in Experiment 1 registered a response on all 32 of the relevant test items (Medium-range in this case), at least 21 of them would need to have harmony in order for the subject to be classified as a successful learner, such that their data would be retained for the present analysis. In many cases subjects did not respond within the allotted three second window, and therefore did not have a registered response for all 32 trials. A subject who only registered 31 responses would have to have chosen 20 of the relevant test items with harmony to surpass the threshold of learning. Subjects who registered either 29 or 30 responses needed to reach 19 responses with harmony. Data from subjects who did not register at least 29 responses (on the 32 relevant test trials) were not considered for this analysis.

For reference, in Experiment 1, eight of the sixteen M-Harm subjects achieved the threshold for learning, having responded with harmony in a sufficient number of

[3] A one-tailed test was used (as opposed to a two-tailed test) since subjects who learned the target pattern are expected to choose the test items adhering to that pattern *more often* than the Control group, rather than simply being significantly different from them.

Medium-range test items, as did six of the sixteen S-Harm subjects, where learning was assessed based on the responses to Short-range test items. In Experiment 2, six of the M-Diss subjects and eleven of the S-Diss subjects were considered successful learners based on the same criteria. Data was collected from further subjects until the number of learners in each of the four groups reached twelve. Of the sixteen subjects in the Control group (the same subjects for both Experiment 1 and 2), no one surpassed the defined threshold at any of the three testing distances. To maintain consistency with the other groups, the analyses below include data only from the first twelve control subjects.

3.2.3.3 Results for Successful Learners in Experiments 1 and 2

After obtaining twelve successful learners in each of the S-Harm, M-Harm, S-Diss, and M-Diss groups, results were analyzed using mixed-effects logistic regressions with structures identical to those used in the above analyses of the first 16 subjects in each condition of Experiments 1 and 2. To highlight the important aspects of the extended analysis, tables summarizing the full regression models are omitted (see Appendix B for the full models), and I present only the odds ratios for each of the harmony and dissimilation conditions, as seen in Table 3.4.[4]

With respect to the two groups that were exposed to harmony, the model indicates that both the M-Harm and S-Harm groups are significantly more likely than the Control group to choose a test item with liquid harmony at *all three* levels of locality given in the test phase of the experiment. This is a deviation from the otherwise consistent result of previous experiments looking at the learning and generalization of consonant harmony locality (Section 2.2.2 of this dissertation; Finley, 2011, 2012; McMullin, 2013; McMullin and Hansson, 2014). I point out that there is a dramatic difference in the size of these odds ratios (maximum of 14.59 for S-Harm at Short-range, minimum of 1.63 for S-Harm at Long-range), but will postpone my discussion of the implications of this result until the next section.

The results obtained for the two dissimilation groups, shown in the bottom two rows of Table 3.4, are more in line with the expected outcome based on the cross-

[4]The results are presented in the same table in order to provide an easy visual comparison of the overall results. Data were analyzed with two separate mixed-effects logistic regressions—one for the harmony conditions and one for the dissimilation conditions.

Table 3.4: Odds ratios comparing 12 successful learners from experimental groups to first 12 control subjects for choosing the target pattern with each of the three testing distances as model baselines for each of the harmony and dissimilation conditions. Contexts encountered in training are in boldface and all cells that reach significance are shaded.

	Type of test item (trigger-target distance)		
	Short-range (cvcvLv-Lv)	Medium-range (cvLvcv-Lv)	Long-range (Lvcvcv-Lv)
M-Harm vs. Control	11.95 (p < 0.001)	**12.94 (p < 0.001)**	3.64 p ≈ 0.007
S-Harm vs. Control	**14.59 (p < 0.001)**	1.78 p ≈ 0.012	1.63 p ≈ 0.031
M-Diss vs. Control	8.86 (p < 0.001)	**14.65 (p < 0.001)**	2.06 p ≈ 0.009
S-Diss vs. Control	**12.01 (p < 0.001)**	1.55 p ≈ 0.106	0.87 p ≈ 0.614

linguistic typology of locality. Subjects in the M-Diss group were significantly more likely than the Control group to choose a test item with two different liquids at all three testing distances. Subjects in the S-Diss group were significantly more likely to do so only at the Short-range distance that they were exposed to in training. S-Diss subjects were just 1.55 times more likely to choose dissimilation in Medium-range test items (did not reach a significance level of $p < 0.05$), and at Long-range they were about equally as likely as the Control group to choose dissimilation (OR = 0.87, $p \approx 0.614$).

3.2.3.4 Discussion: Successful Learners in Experiments 1 and 2

As described above, the results of the first twelve learners in the harmony conditions did not meet expectations in that not only the M-Harm group, but also the S-Harm group, were significantly more likely to choose test items with liquid harmony at all three levels of locality. This deviates from previous experiments as the statistics, strictly speaking, support the conclusion that S-Harm subjects interpreted their relatively local cvcvLv-Lv training as being representative of a pattern with unbounded locality. This result brings up a number of interesting points for discussion.

First, even when the statistics are interpreted in this binary fashion, drawing conclusions based only on whether or not an effect reached a conventional significance level of $p < 0.05$ (as was the original intention for this approach), the result that both sets of subjects applied the pattern with unbounded locality does not contradict the idea that there is a relationship between the types of patterns found cross-linguistically and the types of patterns that are included in the human learner's hypothesis space. For the M-Harm group, the only attested type of locality that their cvLvcv-Lv training items were compatible with was an unbounded pattern, since a dependency across two vowels (and an intervening consonant) implies that the same dependency should hold both at shorter and longer distances. Any other result would contradict the attested typology of locality for patterns of long-distance consonant agreement. However, the cvcvLv-Lv dependency that was presented in the S-Harm training phase is compatible with two different types of locality that are both attested—transvocalic and unbounded. From the preliminary analysis of Experiment 1, which used the first 16 subjects whether they were successful learners or not, I concluded that subjects interpreted the impoverished input conservatively and did not extend it to either of the two greater testing distances. The present analysis, however, shows that when restricting the data to subjects who learned a pattern at the Short-range distance, they do, when taken as a group, tend to generalize the dependency to pairs of liquids that are farther apart. Importantly, this is true not only for the Medium-range, but also for the Long-range test items, in spite of an established reluctance for experimental participants to generalize phonological alternations into initial syllables (Becker et al., 2012, an effect also seen in Finley 2012). As such, when this statistical analysis is interpreted with a strict criterion for 'significant or not', the result does differ from past experiments that were not restricted only to the learners of a pattern, but arguably strengthens the evidence for a strong relationship between learnability and typology.

Further information about the behaviour of the subjects in the harmony conditions can be drawn by considering the relative values of the odds ratios (i.e. the effect size) in each of the cells in Table 3.4, as opposed to only the shading that indicates a p-value of less than 0.05. In particular, note that the S-Harm group is more than fourteen times (OR = 14.59) more likely than the Control group to choose liquid harmony at the Short-range distance that corresponds to their training. This is not at all surprising given that each of the S-Harm subjects was selected for the analysis precisely because they provided a large proportion of harmony responses at this distance. Comparatively, however, the odds ratios at Medium- and Long-range are much smaller at 1.78 and 1.63 respectively. This means that although the effect at each testing distance reached significance, subjects in the S-Harm group were not even twice as likely as the Control group to choose harmony at either of the two distances not encountered in training. The M-Harm subjects, who were about thirteen times (OR = 12.94) more likely than the Control group to choose harmony at the distance they were exposed to in training (Medium-range), also had a large effect for the Short-range distance (OR = 11.95, which indicates a strong tendency to generalize the pattern in to shorter distances) and the odds ratio for Long-range test items was relatively large as well (OR = 3.64). Though the latter figure is perhaps not as impressive as those with OR > 10, it nonetheless provides evidence that subjects in the M-Harm group tend to generalize the pattern *outwards* to word-initial Long-range contexts even more than S-Harm subjects applied the pattern to Medium-range test items.

As a final point of discussion for the results of the learners-only analysis of the harmony conditions, I consider the issue of where non-adjacent dependencies arise in the first place. Evidence suggests that unbounded consonant harmony emerges diachronically from systems that restrict the phonotactic dependency to transvocalic contexts (Dolbey and Hansson, 1999; Gunnar Ólafur Hansson, pers. comm.). In probabilistic terms, the overall tendency for language learners in this experiment—and conceivably for learners of a natural language with transvocalic consonant harmony—is to apply the pattern with a high probability in $\ldots\underline{\mathrm{C}}\mathrm{v}\underline{\mathrm{C}}\ldots$ contexts, and to overextend the pattern (albeit with a much lower probability) into $\ldots\underline{\mathrm{C}}\ldots\mathrm{c}\ldots\underline{\mathrm{C}}\ldots$ contexts. A small effect of this type would likely not contest the stability of a pattern with transvocalic locality, but over time some such patterns could be interpreted as unbounded by a new generation of speakers due to any number of factors (e.g. the number of learners that over-generalized the pattern or the number of lexical items that adhered to an unbounded pattern surpassed some threshold).

I turn now to the results for the two groups in the dissimilation condition, which had provided the original motivation for looking only at the learners of the pattern. Before limiting the analysis in this way, the results of the statistical model were troubling in that they did not support any of the predictions laid out in Section 3.2.1. However, as illustrated in the bottom two rows of Table 3.4, the results now provide evidence that humans learn and generalize patterns of long-distance dissimilation in exactly the same way that they do for harmony—the M-Diss group acquires an unbounded pattern, generalizing to all levels of locality (though, as with harmony, less so to the word-initial context of the Long-range test items), and the S-Diss group interprets the pattern as having strictly transvocalic locality and does not over-generalize to greater distances. I note, however, that there is also a small effect for the S-Diss group when choosing dissimilation in Medium-range test items (OR=1.55). Though the effect does not quite reach statistical significance ($p \approx 0.106$), it nonetheless leaves open the possibility that a diachronic shift from transvocalic to unbounded dissimilation might be predicated by the tendency for some learners to overextend the pattern.

Overall, after restricting the statistical analysis to data from those who surpassed a threshold for learning, the results of these experiments do not support the predictions of the ABC model. Specifically, subjects in the the S-Diss training group would be expected to learn an unbounded pattern, or at least be more likely than subjects in the S-Harm group to generalize to greater distances, but there is no evidence to support either of these two predictions. The observed outcome thus contradicts the ABC model since a transvocalic-only pattern of dissimilation cannot be generated without the addition of redundant constraints that permit further undesirable pathologies (e.g. CORR constraints that stipulate a CVC window for assessing violations; see Section 3.1.2.3 above). Moreover, the M-Diss group showed no tendency to interpret the dissimilation in cvLvcv-Lv contexts as the sort of "beyond-transvocalic" dependency that the ABC model predicts to be possible.

3.3 Summary and Conclusions

This chapter first demonstrated that the Agreement by Correspondence (ABC) model of long-distance consonant interactions (Walker, 2000a,c; Hansson, 2001, 2010a; Rose and Walker, 2004; Bennett, 2013) produces a number of questionable predictions when considering complex instances of consonant harmony and patterns of long-distance consonant dissimilation (Bennett, 2013, 2015), which are not supported by the typology. However, it is important to note that while these pathologies were generated under Bennett's (2013) definition of the correspondence relation (i.e. an equivalence relation that is symmetric, reflexive, and transitive), prior formulations of ABC (e.g. Hansson, 2001, 2010a; Rose and Walker, 2004) do not make the same assumptions. While each proposal leads to slightly different sets of predictions, no version of ABC avoids all of the problematic predictions that were presented above.

The second portion of this chapter presented experimental results that indicate the same patterns of generalization for liquid dissimilation as were previously established for liquid harmony. Learners exposed to dissimilation at Short-range tend not to extend this pattern to further distances. More importantly, learners do elevate an observed Medium-range-only dissimilation pattern to an unbounded dependency, counter to the predictions of the ABC model. The fact that the same learning bias is evidenced for both harmony and dissimilation argues against the way locality relations are referenced in correspondence-based analyses of consonant harmony and dissimilation and weakens the case for Bennett's (2013) "mismatch prediction" regarding these two types of phenomena.

Chapter 4

Locality in Formal Language Theory: A Tier-Based Solution

This chapter seeks to reconcile the apparent shortcomings of the Agreement by Correspondence (ABC; Walker, 2000a,c; Hansson, 2001, 2010a; Rose and Walker, 2004; Bennett, 2013) approach to long-distance consonant interactions by proposing a formal-language-theoretic account of the observed properties of the typology and learnability of such patterns. In approaching phonotactic dependencies from the perspective of formal language theory, the goal is to characterize the set of observed patterns as a class of stringsets (i.e. formal languages; see Section 1.4.2). Strings of segments that adhere to a set of phonotactic restrictions will be grammatical words in the language (stringset), but any word that violates the phonotactics will be ungrammatical (and therefore not a member of the stringset).

As discussed in Section 1.4.2, Johnson (1972) and Kaplan and Kay (1994) establish that any phonological mapping that can be generated with an ordered set of rewrite rules (i.e. A → B / C__D) belongs to the class of regular relations. Furthermore, this means that all stringsets generated by these relations (the surface phonotactics) are members of the regular region of the Chomsky hierarchy (Chomsky, 1956; Rabin and Scott, 1959; see Figure 1.2). The result is that virtually all attested phonotactic patterns are indeed regular, including long-distance consonant agreement and disagreement (Heinz, 2010; Heinz et al., 2011; Payne, 2014). While there is reason to be skeptical of the idea that the full class of regular languages is a plausible definition of what constitutes a possible (and human-learnable) phonotactic dependency (see, e.g., Heinz, 2007, 2010; Lai, 2012), the region can be further broken down into a number of subregular language classes (McNaughton and Papert, 1971; Rogers et al., 2010; Heinz et al., 2011; Rogers and Pullum, 2011). These can be organized into a *subregular hierarchy*, which is shown in Figure 4.1 (repeated from Figure 1.3 in Chapter 1).

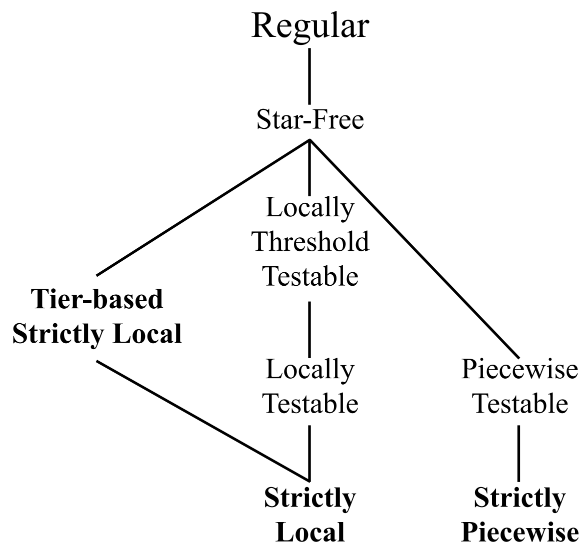

Figure 4.1: Illustration of the subregular hierarchy. The largest class of formal languages (i.e. Regular) is presented on the top, and subset classes are presented below. Each language class is thus a proper subset of any class that is above it and connected by a line. Subregular language classes that are most relevant to this dissertation are presented in boldface.

This chapter provides a foundation for studying the formal properties of long-distance phonotactics within the subregular hierarchy, outlining two alternatives for a characterization of the dichotomy of unbounded vs. transvocalic harmony. I begin by summarizing a modular approach advocated by McMullin and Hansson (2014), who argue that the typology of consonant harmony locality is a result of two distinct learning modules whose combination results in two distinct types of patterns: transvocalic dependencies between consonants may be acquired as Strictly 3-Local (SL_3; McNaughton and Papert, 1971) languages that ban certain …$\underline{\text{C}}$v$\underline{\text{C}}$… sequences while unbounded patterns can instead be learned as Strictly

2-Piecewise (SP_2; Heinz, 2010; Rogers et al., 2010) languages with restrictions on certain …$\underline{C}$…$\underline{C}$… subsequences (Section 4.1). The resulting definition of the learner's hypothesis space under this approach is a union of two subregular classes of stringsets: $SL_3 \cup SP_2$.

While the modular account provides a relatively close fit to the empirical data, I argue in Section 4.2 that it is preferable to characterize long-distance dependencies as Tier-based Strictly 2-Local languages, based on cases of harmony and dissimilation with *blocking* that are easily captured within the TSL_2 region (but not within the $SL_3 \cup SP_2$ region). Furthermore, Section 4.3 demonstrates that each of the unattested patterns (including the ABC pathologies discussed in Section 3.1) cannot be characterized as a TSL_2 stringset. Section 4.4 presents the results of Experiments 3 and 4, which show that patterns *outside of* the TSL_2 region are extremely difficult for humans to learn in the lab—very few subjects who are exposed to such patterns are able to reproduce them in testing, and several subjects seem to learn the dependency as a TSL_2 restriction in spite of overt counter-evidence in the training data against such an interpretation. Section 4.5 summarizes the arguments presented in this chapter and concludes.

As a final note, when characterizing the phonotactics of a language as a stringset, I assume that the patterns are surface true, in that all grammatical words are members of the stringset and any ungrammatical word is not. This stands in contrast to the notion that co-occurrence restrictions are violable constraints. While the present chapter approaches long-distance consonant interactions categorically (i.e. as grammatical or not), I will argue in Chapter 5 that defining constraints as individual subregular stringsets allows for a simple account of several attested patterns that are otherwise relatively complex.

4.1 Strictly Local and Strictly Piecewise Languages

Characterizing a co-occurrence restriction as a member of a subregular class is relatively simple when the dependency is bounded in terms of locality. Such patterns can always be defined as Strictly k-Local (SL_k), where k is the greatest number of segments over which the restriction must be enforced (including those involved in the restriction). Substrings of length k are called k-factors (i.e. segment n-grams), and the grammar for an SL_k language can be thought of as a list of permitted (or, equivalently, prohibited) k-factors. For example, we can define a language that allows only CV syllable structure as SL_2, since the restriction targets pairs of adjacent segments (2-factors, bigrams). With a simplified alphabet of $\Sigma = \{\mathrm{C}, \mathrm{V}\}$, the grammar for such a language needs only to include a few 2-factor restrictions: $G = \{$*CC, *VV, *#V, *C#$\}$, where # is a word boundary.[1] Transvocalic variants of consonant harmony and dissimilation can also be defined as SL languages, though k must provide a slightly larger window for application of the co-occurrence restriction. As defined in Chapter 2, transvocalic harmony does not hold over intervening consonants, but does hold across intervening vowels in …CvC… sequences. Consider a simplified version of the Koyra pattern shown in (1) (reproduced from above; see Section 2.1 for full description of the data), which reduces the segment inventory to $\Sigma = \{$s, ʃ, t, a$\}$, where ‘s’ represents a [+anterior] sibilant, ‘ʃ’ is a [–anterior] sibilant, ‘t’ is any other consonant, and ‘a’ is any vowel. The grammar would simply be a list of k-factors (with $k \leq 3$) not permitted in the language[2], namely any that include both [s] and [ʃ]: $G = \{$*ʃas, *saʃ, *sʃ, *ʃs$\}$.[3]

(1) Transvocalic sibilant harmony in Koyra (Koorete; Hayward, 1982)

a.	/tim-d-osːo/	tindo**s**ːo	‘he got wet’	
b.	/patʃ-d-osːo/	pa**tʃ**ː**oʃ**ːo	‘it became less’	*pa**tʃ**ː**os**ːo
c.	/giːʒ-d-osːo/	giː**ʒ**ː**oʃ**ːo	‘it suppurated’	*giː**ʒ**ː**os**ːo
d.	/ʃod-d-osːo/	ʃodːosːo	‘he uprooted’	
e.	/ʔatʃ-ut-d-osːo/	ʔatʃutːosːo	‘he (polite) reaped’	

For patterns with unbounded locality, however, one of the main challenges is that any formal language characterization must allow for, in principle, an infinite num-

[1] In general, this dissertation sets aside the issue of word boundaries. Technically speaking, they are usually incorporated into the theory by defining the language with a set of symbols Σ that is augmented with two word-boundary symbols—one for the beginning of a word and one for the end.

[2] 2-factors are included in SL_3 grammars in order to reduce the number of k-factors that must be listed. A full grammar for this example would also include 3-factors such as $\{$*ʃʃs, *stʃ, *asʃ, …$\}$.

[3] The data in (1) only demonstrate the ungrammaticality of *[–ant]V[+ant]. However, the restriction holds more generally as a morpheme structure constraint, such that no well-formed root may contain *[–ant]V[+ant] or *[+ant]V[–ant]. Also, since the language has no suffixes that contain an underlying [–ant] sibilant, there is no evidence that, e.g., *saʃ should actually be permitted (Hayward, 1982; Hansson, 2010a).

ber of non-participating segments to intervene between the relevant pair. Unbounded dependencies therefore cannot be SL_k for any value of k, since the restrictions hold at all distances, including $k+1$.

Strictly k-Piecewise (SP_k) languages are those that can be defined in terms of linear precedence relations. Using *subsequences* of length k, which give information about linear order without reference to distance or intervening material, it is possible to capture the less restricted nature of unbounded consonant harmony and dissimilation. For example, the 2-subsequences (precedence relations) of a word [sataʃa] would include {s…a, s…t, s…ʃ, a…t, a…a, a…ʃ, t…a, t…ʃ, ʃ…a}. Since SP languages encode information about the order of segments while ignoring distance, most attested cases of unbounded consonant harmony can be characterized as SP_2 (including patterns with asymmetric directionality or feature dominance; Heinz, 2010). This includes the case of Aari sibilant harmony shown in (2), in which the perfective suffix /-s/ surfaces as [-ʃ] when it is preceded by a [–ant] sibilant at any distance (see Section 2.1 for full description of the data). Using a simplified segment inventory, as above, with $\Sigma = \{$s, ʃ, t, a$\}$, this pattern of sibilant harmony with unbounded locality can be characterized as SP_2 with the following grammar: $G = \{$*ʃ…s, *s…ʃ$\}$.

(2) Unbounded sibilant harmony in Aari (Hayward, 1990)

a.	/baʔ-s-e/	ba**ʔs**e	‘he brought’
b.	/ʔuʃ-s-it/	ʔu**ʃʃ**it	‘I cooked’
c.	/tʃ’a̤ːq-s-it/	**tʃ’**a̤ːq**ʃ**it	‘I swore’
d.	/ʃed-er-s-it/	**ʃ**eder**ʃ**it	‘I was seen’

Finally, it is also important to note that transvocalic dependencies cannot be characterized as SP_2, since precedence relations offer no information about the distance between two segments. If the target pattern is transvocalic sibilant harmony, the grammaticality of a word such as [sadaʃ] (which does not include any …S̲vS̲… substrings) implies that any word with a s…ʃ subsequence—including a word such as [saʃ], which violates transvocalic harmony—should also be grammatical (barring, of course, any other phonotactic violations). In fact, this holds more generally, in that any phonotactic dependency that is bounded by some measure of locality is not SP_k for any value of k (provided, of course, that the value of k does not exceed the longest word in the language).

4.1.1 Learning Bias and the Argument for Modular Learning

Given some upper bound on k, sound patterns that are SL_k or SP_k are proven to be efficiently learnable in the limit from positive evidence (Gold, 1967) by relatively simple learning algorithms. For SL_k patterns, the learner simply keeps track of all n-grams, or k-factors, encountered in the training data (Garcia et al., 1990; Heinz, 2007). For SP_2 patterns, the learner instead records all encountered k-subsequences (a precedence learner; Heinz, 2010). Based on the fact that the transvocalic and unbounded dependencies found in natural language instantiate well-defined but distinct formal classes of languages, McMullin and Hansson (2014) argue that phonotactic learning is *modular*, with different learning algorithms responsible for detecting different types of phonotactic regularities (Heinz 2010; for arguments for modular learning of phonological vs. syntactic patterns, see Heinz and Idsardi 2011, Lai 2012). Both types of locality can be efficiently learned under the assumption that the phonological learner contains at least two sub-modules—an n-gram learner for transvocalic harmony (SL_3) and a precedence learner for unbounded harmony (SP_2). By contrast, the unattested locality patterns, such as those depicted in Table 2.1, are neither SL_k nor SP_k (at least not for any reasonable value of k). From this perspective, the gaps in the typology of locality relations in consonant harmony are thus a direct reflection of learning bias. Alternative locality patterns, which are conceivable but unattested, are situated outside of the learner's hypothesis space and are inaccessible (diachronically and synchronically) due to this inductive bias that operates in phonotactic learning. Such patterns are beyond the capabilities of either the SL or the SP module and will therefore not be acquired and replicated faithfully by learners, other things being equal (as seen in the results of Experiments 1 and 2; see also Section 4.4 below for results of Experiments 3 and 4).

4.1.2 Evidence Against this Approach from Blocking

Though Heinz (2010) demonstrates that nearly all cases of unbounded consonant harmony from two typological surveys that were available at the time (Hansson, 2001; Rose and Walker, 2004) can be described as members of the SP_2 class of formal languages, the potential extension of this approach to all types of long-distance dependencies is impeded by a number of attested phonotactic patterns that are not SP_2 (nor SL_k for that matter). For example, Heinz (2010) himself notes that certain cases of consonant dissimilation are known to exhibit segmental blocking effects (Odden, 1994; Heinz et al., 2011; Bennett, 2013). In the case of Georgian liquid dissimilation, illustrated in (3) with the liquids presented in boldface, the adjectival suffix /-uri/ surfaces as [-uli] when preceded by [r] anywhere in the word, except if [l] intervenes.[4]

(3) Liquid dissimilation with blocking in Georgian
(Fallon, 1993; Odden, 1994; Bennett, 2013)

a. dan-u**r**i 'Danish'
b. p'o**l**on-u**r**i 'Polish'
c. ung**r**-u**l**i 'Hungarian'
d. ap**r**ik'-u**l**i 'African'
e. avst'**r**a**l**-u**r**i 'Australian'
f. ka**r**t**l**-u**r**i 'Kartvelian'
g. bu**l**ga**r**-u**l**i 'Bulgarian'

In the above data, (3a)-(3b) show that the suffix /-uri/ surfaces faithfully as [-uri] when the root contains no liquids, or when the root contains an [l]. As seen in (3c)-(3d), when the root contains an [r] at any distance away from the suffix, /-uri/ instead surfaces as [-uli]. However, (3e)-(3f) demonstrate a segmental blocking effect—when [l] intervenes between an [r] in the root and the suffix, it surfaces faithfully as [-uri] in spite of the presence of a preceding [r] in the root. Finally, (3g) shows that the pattern is blocked only when the [l] *intervenes* between two [r]s, not anytime there is an [l] in the root.

Note that the pattern found in Georgian cannot be classified as SP_2 with a restriction against *r…r subsequences. Such an analysis would account for the basic generalization seen in (3a)-(3d), but would fail to permit cases like (3e)-(3f),

[4]Fallon (1993) provides evidence that the /r/→[l] alternation does not occur only for the adjectival suffix /-uri/ (which, when combined with the name of a country, denotes the nationality of a thing, not a person), but also for a number of other suffixes containing /r/ that adhere to the same generalization (i.e. triggered by a preceding [r], blocked by intervening [l]).

which exhibit a blocking effect, as these still contain a supposedly-banned *r…r subsequence. Likewise, because of words like (3g), it cannot be analyzed with a relative ranking of two constraints, each defined as an SP_2 stringset, such as *[l…l]$\gg$*[r…r] (cf. *X…X constraints discussed by Pulleyblank, 2002).

The typological evidence thus suggests that a description of long-distance phonotactics in terms of precedence relations is too restrictive, accounting for only a subset of attested patterns. The remainder of this chapter pursues a different strategy, demonstrating that the Tier-based Strictly 2-Local class of formal languages (Heinz et al., 2011) not only encompasses each of the attested parameters of locality and blocking—some of which the ABC model cannot easily handle—but also excludes the unattested patterns presented in Table 2.1 as well as the pathological patterns predicted by ABC, as described in Section 3.1.

4.2 Tier-Based Strictly 2-Local Languages

The precedence relations encoded in an SP_2 grammar provide a convenient solution for ignoring irrelevant intervening material. As an alternative means of achieving the same goal, the Georgian pattern shown in (3) can instead be thought of as a restriction against contiguous segment pairs {*ll, *rr}, where adjacency is crucially assessed only among liquid consonants.[5] Patterns that can be similarly described are members of the Tier-based Strictly 2-Local class of formal languages (TSL_2; Heinz et al., 2011). This characterization of long-distance dependencies, while still relatively simple, can account for several typological properties of locality that are presented throughout the remainder of this section.

In more formal terms, TSL_k languages can be defined as follows. For some alphabet Σ (a segment inventory), the grammar G of a Tier-based Strictly k-Local language is a two-tuple $G = \langle T, S \rangle$, where the tier T is some subset of Σ over which adjacency is assessed, and S is the set of k-factors permitted on that tier (for an exhaustive formal definition of TSL languages and proofs for several computational

[5] The inclusion of *ll is a slight oversimplification, in order to provide a generalization that is directly comparable the pattern used in Experiment 2. While the data available in Fallon (1993) do not include any root-internal [l…l] subsequences (unless an [r] intervenes, as in [liberal-ur-] ‘liberal’), there is another suffix /-eli/, which denotes the nationality of a person, and never surfaces as *[-eri], even in cases such as [p’olon-eli] ‘Polish’.

properties of the TSL class, see Heinz et al., 2011; Jardine and Heinz, 2015). As shown in Table 4.1, a tier can be a set of segments that corresponds to a natural class, but since it is mathematically defined simply as a subset of the segment inventory, it could also be any arbitrary collection of segments.

Table 4.1: Example tier-based strings for a hypothetical word [pireʃaʃolus], when $\Sigma = \{$p, s, ʃ, r, l, i, e, a, o, u$\}$

Contents of *T*	Description of *T*	Tier-based string
Σ	all segments	**pireʃaʃolus**
{p, s, ʃ, r, l}	consonants	**p**i**r**e**ʃ**a**ʃ**o**l**u**s**
{i, e, a, o, u}	vowels	p**i**r**e**ʃ**a**ʃ**o**l**u**s
{s, ʃ}	sibilants	pire**ʃ**a**ʃ**olu**s**
{r, l}	liquids	pi**r**eʃaʃo**l**us
{p, ʃ, i, u}	arbitrary set	**pi**re**ʃ**a**ʃ**ol**u**s

Note that to arrive at the tier-based string, Heinz et al. (2011) define an erasing function that removes any segments that are not in *T*.[6] Following Jardine and Heinz (2015), I use *R* to denote the set of tier-based 2-factor restrictions (i.e. the complement of *S* with respect to all possible tier-based 2-factors), as it is often more convenient to describe a TSL_2 grammar as $G = \langle T, R \rangle$. For the Georgian liquid dissimilation pattern presented in (3), the components of the TSL_2 grammar encoding the dependency are $T = \{$l, r$\}$ and $R = \{$*ll, *rr$\}$.

4.2.1 Consonant Harmony from the TSL_2 Perspective

Patterns of long-distance consonant agreement can equally be described in TSL_2 terms. The unbounded sibilant harmony pattern of Aari shown in (2) is a restriction on sequences of [αanterior][–αanterior] sibilants on a tier that includes all and only [+strident] segments. Continuing from the above examples with a simplified inventory of $\Sigma = \{$s, ʃ, t, a$\}$, this pattern can be generated by the following TSL_2 grammar: $G = \langle T = \{$s, ʃ$\}, R = \{$*ʃs, *sʃ$\}\rangle$. Heinz (2010) hesitates to de-

[6] Specifically, $E_T(\sigma_1 \ldots \sigma_n) = u_1 \ldots u_n$, where $u_i = \sigma_i$ iff $\sigma_i \in T$ and $u_i = \lambda$ otherwise, where λ denotes the empty string (Heinz et al., 2011, p. 60)

scribe patterns of long-distance consonant agreement in terms of tiers, citing a lack of known systems that exhibit blocking. I argue however, that patterns with the widely attested locality type of transvocalic consonant harmony, such as the Koyra sibilant harmony shown in (1), can be recast as long-distance dependencies that are *blocked by any intervening consonant*. From the perspective of TSL_2 languages, the phonotactic grammar of Koyra bans sequences of [αant][–αant] sibilants, exactly as in the unbounded case of Aari, but in Koyra violations are assessed on a tier that includes all of the consonants, rather than only sibilants. Forms like (1d) [ʃodːosːo] 'he uprooted' are grammatical precisely because another consonant, in this case [d], remains present when the segment string is reduced to the consonants. The [d], intervening between a pair of sibilants that constitutes a member of *R*, can therefore be construed as a 'blocker', since it interrupts a sequence of sibilants that would otherwise violate the phonotactic grammar.

Further motivation for including long-distance consonant agreement within the scope of the TSL_2 approach is that additional cases have come to light that involve more obvious instances of consonant harmony with blocking. Relevant cases include Kinyarwanda (described in Section 5.1.1; Walker and Mpiranya, 2005; Hansson, 2007; Walker et al., 2008), Imdlawn Tashlhiyt Berber (described in Section 5.3.1.2; Elmedlaoui, 1995; Hansson, 2010b), and Slovenian (Jurgec, 2011), which is shown below in (4) with all coronal obstruents in boldface.

(4) Sibilant harmony with blocking in Slovenian (Jurgec, 2011)

a.	**s**pi	'sleeps'	**ʃ**pi-**ʃ**	'(you) sleep'
b.	**z**a-klɔn	'shelter'	ʒa-klɔn-i**ʃtʃ**e	'bomb shelter'
c.	**ts**epə**ts**	'fool'	**tʃ**ep**tʃ**-ək	'fool-DIM'
d.	**s**i**t**	'full'	na-**s**i**t**-i**ʃ**	'(you) feed'
e.	**z**i**d**a	'(s/he) builds'	**z**i**d**a-**ʃ**	'(you) build'

The data in (4) demonstrate that the regressive sibilant harmony pattern found in Slovenian, which bans *[+ant]…[–ant] subsequences in (4a)-(4c), is blocked when a coronal obstruent such as [t] or [d] intervenes, as in (4d)-(4e). Note that in (4b) the coronal sonorants [n] and [l] are transparent, just like non-coronals are. As such, the grammar for Slovenian would prohibit 2-factors of, e.g., $R = \{$*sʃ, *zʃ, *t͡st͡ʃ, etc$\}$, but the relevant tier would include all coronal obstruents (as opposed to just the sibilants, or all of the consonants).

4.2.2 Locality as a Consequence of the Tier

In terms of TSL_2 grammars, the distinction between unbounded dependencies, dependencies with blocking, and transvocalic dependencies can thus be attributed to a difference in the particular subset of Σ that comprises the designated tier T (and its relationship to R), rather than any change to R itself. The grammars presented in Table 4.2 show this for three hypothetical languages with sibilant harmony, which are representative of the range of attested patterns (e.g. Aari, Slovenian, and Koyra, respectively). In order to facilitate a direct comparison between each type, the segment inventory is restricted to $\Sigma = \{\text{s, ʃ, p, t, a}\}$, and the set of prohibited 2-factors is always $R = \{\text{*sʃ, *ʃs}\}$.

Table 4.2: TSL_2 grammars for three types of sibilant harmony

Type of pattern	Σ	T	R
Unbounded	{s, ʃ, p, t, a}	{s, ʃ}	{*sʃ, *ʃs}
Blocking	{s, ʃ, p, t, a}	{s, ʃ, t}	{*sʃ, *ʃs}
Transvocalic	{s, ʃ, p, t, a}	{s, ʃ, p, t}	{*sʃ, *ʃs}

With Σ and R held constant, Table 4.2 shows that variation in (what appears to be) locality is merely a by-product of manipulating the contents of the relevant tier T over which violations are assessed. Unbounded dependencies are a result of a tier that includes only segments that are present in members of R. If any additional segments are included in T, such as a coronal obstruent [t], these will block the pattern. The transvocalic locality type arises when all other consonants are also included in T (and are hence blockers), but the class of vowel segments is systematically absent from the tier.

More generally, the set of segments occurring in the members of R (such as {s, ʃ} for the languages above) can be thought of as a set of *harmonic* segments (i.e. potential triggers or targets), which I will denote P (where each $\sigma_x \in \Sigma$ is a member of P if and only if there is some 2-factor $\sigma_x\sigma_y$ or $\sigma_y\sigma_x$ that is in R). Note that P is a subset of T, and likewise, T is by definition a subset of Σ. Because of this property (i.e. $P \subseteq T \subseteq \Sigma$), there are overall just three possible types of segments. First, any segment that occurs in P (and is therefore in both T and Σ) will participate in the dependency as a potential trigger or as a target for some repair strategy (e.g. harmony or dissimilation). Second, any segment that is in T (and therefore also in Σ), but is not present in P, will act as a neutral segment that blocks interaction between two segments on either side of it. Finally, a segment that is in Σ but not T (and therefore not in P) will be neutral and transparent to the dependency. This is summarized in Table 4.3, for some segment $\sigma \in \Sigma$.

Table 4.3: Three types of segments in a TSL_2 grammar

Type of segment	$\sigma \in \Sigma$?	$\sigma \in T$?	$\sigma \in P$?
Harmonic (trigger/target)	✓	✓	✓
Neutral (opaque)	✓	✓	✗
Neutral (transparent)	✓	✗	✗

A final relevant property of the TSL_2 language class is that when $T = \Sigma$, a stringset can also be described as a simple Strictly 2-Local pattern (i.e. restrictions against string-adjacent segments). As a result, the TSL_2 region properly includes all SL_2 stringsets (a result that holds more generally for any value of k; Heinz et al., 2011).

4.3 Pathological Patterns That Are Not TSL_2

I have now shown how the attested types of locality for long-distance consonant agreement and disagreement with and without blocking are easily captured by describing phonotactic patterns as members of the Tier-based Strictly 2-Local class of formal languages. In this section, I argue that unattested variants of locality, as well as the unusual varieties of long-distance dependency patterns that are predicted to be possible by the basic architecture of the ABC model (see Section 3.1) are pathological not merely in terms of typological attestation but also from the standpoint of computational complexity and learnability. Table 4.4 shows certain unattested types of locality that were discussed (with respect to consonant harmony) in Section 2.1, none of which can be described as a TSL_2 pattern.

The patterns in the first two rows of Table 4.4, dependencies that hold across

Table 4.4: Unattested variants of long-distance locality

Locality	…C̲vC̲…	…C̲vcvC̲…	…C̲vcvcvC̲…
$\leq$ 1 consonant	+	+	–
$=$ 1 consonant	–	+	–
$\geq$ 1 consonant	–	+	+

either at most one or exactly one consonant, are still TSL$_k$ languages, but require at least $k = 3$. For an example using $\Sigma = \{$s, ʃ, t, a$\}$, if the target pattern is sibilant harmony that holds across *at most one* or *exactly one* intervening consonant, one must keep track of at least three consonants in order to prohibit words such as *[sataʃa] and *[ʃatasa] while still permitting [satataʃa] and [ʃatatasa]. While both patterns are TSL$_3$ with $T = \{$s, ʃ, t$\}$, the difference between the two locality variants is whether or not the grammar also prohibits words that include tier-based 3-factors $\{$Csʃ, Cʃs, sʃC, ʃsC$\}$, where C is any of $\{$s, ʃ, t$\}$. That is, words such as [saʃata] and [taʃasa] would be ungrammatical if the dependency holds across up to one consonant (as in the first row of Table 4.4), but they would be grammatical for a pattern that holds across exactly one consonant (the second row of Table 4.4).

A dependency that holds across *at least one* consonant (as in the last row of Table 4.4), is not TSL$_k$ for any value of k, as the phonotactic legality of a word cannot be determined solely in terms of presence vs. absence of individual k-factors regardless of how the tier T is construed. Again using sibilant harmony as an example, such a language would need to permit sequences of …saʃa…, …ʃasa…, but any strings including *…sataʃa…, *…ʃatasa… would be prohibited. Crucially, the latter two restrictions cannot be ruled out simply by setting T to include all consonants and having R containing the 3-factors $\{$*stʃ, *ʃts$\}$. This is because the pattern of sibilant harmony is meant to apply regardless of the number of intervening consonants (provided it is more than one), as in *…satataʃa…, *…satatataʃa…, etc. There is thus no upper bound on k that will suffice to rule out all illegal sequences, since a word of the form *sa…(ta)$^{k+1}$…ʃa will always be erroneously classified as grammatical. Instead, a phonotactic pattern of this type falls into (a tier-based instantiation of) the Locally Testable class (Rogers and Pullum, 2011), which is defined in terms of Boolean operations over sets of *k*-factors. In the present example an illegal word is one that contains both a member of {ts, #s} and {tʃ, #ʃ} among its 2-factors on the consonantal tier (augmented by word boundaries).[7]

Recall from Section 3.1.2.2 that the ABC framework can generate phonotactic dependencies of exactly this type, provided the dependency is a pattern of dissimilation (which I call 'beyond-transvocalic' dissimilation). Interestingly, if the pattern applies to *identical* consonants, such as two liquids that are both [+lateral] or both [–lateral] (i.e. a beyond-transvocalic analogue of the Georgian case seen in (3) above), it lies outside even the Locally Testable class. This is because the grammar would need to be able to count the number of instances of certain *k*-factors; in the case of beyond-transvocalic liquid dissimilation with $\Sigma = \{$l, r, t, a$\}$, a word is illegal if it contains *two or more* occurrences of one of the 2-factors in {rt, tr, lt, tl} on the consonant tier. The relevant class is therefore (a tier-based instantiation of) the Locally Threshold Testable languages (Rogers and Pullum, 2011).

The computational status of phonotactic patterns with beyond-transvocalic locality is thus somewhat analogous to the (unattested) "first-last harmony" pattern described by Lai (2012), where words of the structure #s…ʃ# or #ʃ…s# are banned (but both of, e.g., #s…s…s# and #s…ʃ…s# are permitted). Lai's artificial language learning experiments showed a failure to learn first-last harmony, suggesting that Locally Testable patterns that lie outside the TSL (and SP) regions of the subregular hierarchy (Figure 4.1) are beyond the grasp of the human phonological learner—a proposal that is further supported by evidence from Experiments 3 and 4 presented below in Section 4.4.

Finally, neither of the two ABC pathologies discussed in Section 3.1.1 can be characterized as a TSL_2 pattern. The status of the bizarre parity-sensitive harmony pattern generated by high-ranked PROXIMITY (see Section 3.1.1.2 for full description and tableaux) is unclear at present (it may, for instance, be Regular but not Star-Free), but it in any case falls beyond the SL, SP or TSL subregular classes. Also outside those classes are the "agreement by proxy" effects discussed in Section 3.1.1.1. In the example case, where /s…g/ → [z…g] assimilation is dependent on a nearby [x], the sound pattern cannot be expressed in TSL_k terms for any k, even

[7] Note that 'ts' and 'tʃ' do not denote affricates here, but 2-factors of [t]+[s] or [t]+[ʃ].

if the tier T is defined as $\{s, z, g, x\}$. This is because the pattern holds no matter how many additional instances of [g] or [s] intervene between the [s…g] pair and the "proxy" [x].

4.4 Experiments 3 and 4: Rich-Stimulus Training

This chapter has thus far established that we can define the boundaries of a learner's hypothesis space in terms of formal language theory. After first offering two alternatives for doing so within the subregular hierarchy, I argued that the class of Tier-based Strictly 2-Local formal languages provides an accurate approximation of the cross-linguistic properties of long-distance dependencies with respect to both locality and blocking. I now present experimental evidence that the TSL_2 approach is on the right track in terms of offering a level of complexity that accurately defines the hypothesis space of a human learner. Specifically, Experiments 3 and 4 show that very few subjects who are exposed to patterns of liquid harmony or dissimilation that hold at a Medium-range distance, but demonstrably fail to hold at Short-range, are able to learn anything at all from their training. Furthermore, of those subjects who seem to pick up on a CV$\underline{\text{L}}$VCV-$\underline{\text{L}}$V dependency, many of them appear to learn it as a TSL_2 restriction that applies to Short-range contexts as well (i.e. as an unbounded dependency), in spite of the counter-evidence provided in their training.

4.4.1 Motivation for Experiments 3 and 4

Experiments 1 and 2 used a "Poverty-of-Stimulus" paradigm (e.g. Wilson, 2006; Finley and Badecker, 2009) to determine not only whether humans can learn a dependency between non-adjacent liquids, but also whether or not they generalize the learned pattern to contexts that were purposely withheld from them in the training phase. Recall that the results suggested that subjects who learn liquid harmony or dissimilation from Medium-range CV$\underline{\text{L}}$VCV-$\underline{\text{L}}$V items tend to generalize the pattern to all levels of locality. By contrast, subjects who learn from Short-range CVCV$\underline{\text{L}}$V-$\underline{\text{L}}$V items tend to restrict the pattern to the Short-range transvocalic distance (though see Section 3.2.3.3 for evidence that some subjects in the S-Harm group may generalize in an unbounded fashion).

Of relevance to the current discussion is that neither of the two Medium-range training groups in Experiments 1 or 2 (M-Harm or M-Diss) showed evidence of having learned a pattern with beyond-transvocalic locality (i.e. by applying the pattern only to the Medium- and Long-range test items, but not to the Short-range test items). These results were expected based on the typology of locality in long-distance phonotactics. Furthermore, the fact that the M-Diss group in particular did not learn a beyond-transvocalic pattern was used as an argument against the predictions of the Agreement by Correspondence framework, since the factorial typology of ABC predicts that beyond-transvocalic dissimilation should be a possible pattern (while, notably, beyond-transvocalic *harmony* should not).

This chapter has argued that beyond-transvocalic dependencies cannot exist in natural language because they are inaccessible to the human learner, whose hypothesis space for phonotactic patterns is defined by the proposed TSL$_2$ region. Recall that Section 4.3 established that beyond-transvocalic dependencies are outside of this region. However, the results of Experiments 1 and 2 do not necessarily provide conclusive evidence for leaving patterns with beyond-transvocalic locality outside the space of human-learnable languages. Humans may be capable of learning such a dependency, but simply have a preference for the (formally less complex) unbounded version of the pattern when presented with training items that are compatible with both, as was the case in the "Poverty-of-Stimulus" design used in Experiments 1 and 2.

For example, consider the possibility that consonant harmony with either unbounded or beyond-transvocalic locality is a possible pattern, and recall that the M-Harm group was exposed to liquid harmony in Medium-range (CV$\underline{\text{L}}$VCV-$\underline{\text{L}}$V) contexts. Since the training phase presented no information about the behaviour of liquids at the Short-range (CVCV$\underline{\text{L}}$V-$\underline{\text{L}}$V) distance or the Long-range ($\underline{\text{L}}$VCVCV-$\underline{\text{L}}$V) distance, the training data were, in principle, compatible with either unbounded harmony or beyond-transvocalic harmony. Even though subjects in the M-Harm group showed evidence of having generalized the dependency to all distances, this does not mean that they are incapable of learning beyond-transvocalic harmony. Instead, the strongest conclusion we can draw is that, when presented with data that are ambiguous between unbounded and beyond-transvocalic locality, learners prefer the unbounded interpretation.

The purpose of Experiments 3 and 4 is therefore to provide subjects with a “Rich-Stimulus” training phase, offering (conflicting) evidence about what happens to pairs of liquids that co-occur in Short- vs. Medium-range contexts in order to determine whether or not the TSL_2 region is defining a hypothesis space that is too restrictive.

4.4.2 Methodology

4.4.2.1 Participants, Stimuli, and Procedure

Participants were recruited and compensated in the same way as previous experiments, resulting in 32 new participants for both Experiment 3 (26 female, 6 male, mean age 23) and Experiment 4 (21 female, 11 male, mean age 22), which were performed using the same stimulus set and procedures that were used for Experiments 1 and 2 (see Section 2.2.1).

4.4.2.2 Training Conditions

Data was collected for four new groups of subjects who were again tested on the same 96 test items, but differed in the types of words and phonotactic patterns they were exposed to in training. Recall that in Experiments 1 and 2, only 50% of the training stems for each of the S-Harm, M-Harm, S-Diss, and M-Diss conditions contained a liquid (either [l] or [ɹ]) in the relevant position. The remaining half of the stems contained no liquids whatsoever, and so all consonants remained faithful when the suffixes [-li] or [-ɹu] were attached. In Experiments 3 and 4, this second (faithful) half of the training data was replaced with a different set of stems. All segments in the replacement stems continued to stay faithful when the suffixes were added, but in this case they contained a liquid [l] or [ɹ]. Depending on whether the first portion of the training data showed evidence of alternating liquids at the Short- or Medium-range distance, the faithful liquids were located at the *opposite* distance. That is, subjects were given explicit evidence that liquids alternate at one of the Short- or Medium-range distances, but that they *do not alternate* at the other distance. There are thus four new groups of this type, which are labelled S-Harm-M-Faith, M-Harm-S-Faith for Experiment 3 and S-Diss-M-Faith, M-Diss-S-Faith for Experiment 2. These labels follow the conventions established for previous experiments ("S" indicates Short-range CVCVLV stems, "M" indicates Medium-range CVLVCV stems, and "Harm" or "Diss" indicates that the target pattern at this distance was harmony or dissimilation, respectively). To further differentiate them from previous groups, the distance that showed counter-evidence in the form of faithful liquids is indicated by "-S-Faith" or "-M-Faith" at the end of each group's label.

As an illustration of what the subjects were exposed to in training, Table 4.5 provides a breakdown of the number and type of stimuli encountered by the S-Harm-M-Faith group in Experiment 3. Note that a pattern of this type is compatible

Table 4.5: Example of S-Harm-M-Faith training in Experiment 3.

Training triplet	Type and number of items
…begeli…begeɹi-ɹu…begeli-li… …domelo…domelo-li…domeɹo-ɹu…	48 CVCVLV stems with [l] (Short-range harmony)
…mopeɹe…mopeɹe-ɹu…mopele-li… …tetoɹi…tetoli-li…tetoɹi-ɹu…	48 CVCVLV stems with [ɹ] (Short-range harmony)
…pilede…pilede-li…pilede-ɹu… …nelogi…nelogi-ɹu…nelogi-li…	48 CVLVCV stems with [l] (Medium-range faithfulness)
…koɹupe…koɹupe-li…koɹupe-ɹu… …guɹoto…guɹoto-ɹu…guɹoto-li…	48 CVLVCV stems with [ɹ] (Medium-range faithfulness)

with the (attested) transvocalic variant of consonant harmony and is predicted to be learned as is. While it may seem strange to include a group like this, given that the results of Experiment 1 show that the S-Harm group is able to learn a transvocalic pattern even without being exposed to faithful liquids at the Medium- or Long-range distance, the S-Harm-M-Faith group (and the S-Diss-M-Faith group) is important for the interpretability of results and for a more direct comparison with Experiments 1 and 2. Consider the possibility that the responses given by the M-Harm-S-Faith group are not statistically different from the Control group. We cannot necessarily attribute a failure of learning to the fact that the pattern was more complex than those in Experiments 1 and 2, since there are methodological changes. For example, every stem that subjects in Experiment 3 (and Experiment 4) were exposed to in training contained a liquid, and this difference alone might lead to more difficulty in processing the input and learning a pattern from it. It is therefore useful to have not only a comparison group in the form of the Control subjects, but also in the S-Harm-M-Faith group (and S-Diss-M-Faith) that allows for an overall comparability with the previous experimental findings.

Table 4.6 provides examples from the training phase of the M-Diss-S-Faith group in Experiment 4. This pattern complies with the (unattested and formally complex) beyond-transvocalic variant of dissimilation and is therefore predicted by the TSL_2 hypothesis to be inaccessible to human language learners, even though the factorial typology of ABC predicts that it is a possible pattern (see Section 3.2.1).[8]

Table 4.6: Example of M-Diss-S-Faith training in Experiment 4.

Training triplet	Type and number of items
…begeli…begeli-ɹu…begeli-li…	48 CVCVLV stems with [l]
…domelo…domelo-li…domelo-ɹu…	(Short-range faithfulness)
…mopeɹe…mopeɹe-ɹu…mopeɹe-li…	48 CVCVLV stems with [ɹ]
…tetoɹi…tetoɹi-li…tetoɹi-ɹu…	(Short-range faithfulness)
…pilede…piɹede-li…pilede-ɹu…	48 CVLVCV stems with [l]
…nelogi…nelogi-ɹu…neɹogi-li…	(Medium-range dissimilation)
…koɹupe…koɹupe-li…kolupe-ɹu…	48 CVLVCV stems with [ɹ]
…guɹoto…guloto-ɹu…guɹoto-li…	(Medium-range dissimilation)

4.4.3 Results and Analysis

Data was first collected for 16 native English speakers in each of the four new experimental groups, and was analyzed in much the same way as was done for Experiments 1 and 2 (see Section 2.2.2.3 for a full description), using the same group of control subjects (who were not exposed to any stems with liquids in the training phase) for a baseline comparison in the mixed-effects logistic regression

[8] Note that the Agreement by Correspondence framework only predicts the possibility of a strictly beyond-transvocalic variant of locality for patterns of consonant *disagreement*, and there is no way to generate a beyond-transvocalic version of consonant harmony using ABC constraints.

analyses. Note that the results presented below omit the full statistical models (see Appendix B for full summaries) in favour of the more informative tables of odds ratios. The OR values in Table 4.7 are presented as an indication of how each of the groups performed *as a whole*. Plots comparing each group to the Control group at each testing distance are provided below in Section 4.4.3.1 (for Experiment 3) and Section 4.4.3.2 (for Experiment 4), and results are discussed in more detail there, especially with respect to individual subject performance.

Table 4.7: Odds ratios comparing groups in Experiments 3 and 4 to Control subjects for choosing the target pattern (values extracted from mixed-logit models relevelled at each testing distance). Contexts encountered in training are in boldface and all cells that reach significance are shaded.

	Type of test item (trigger-target distance)		
	Short-range (cvcvLv-Lv)	Medium-range (cvLvcv-Lv)	Long-range (Lvcvcv-Lv)
M-Harm-S-Faith vs. Control (Experiment 3)	2.22 ($p < 0.001$)	**2.50 ($p < 0.001$)**	1.74 ($p \approx 0.011$)
S-Harm-M-Faith vs. Control (Experiment 3)	**7.42 ($p < 0.001$)**	1.57 ($p \approx 0.042$)	1.50 ($p \approx 0.062$)
M-Diss-S-Faith vs. Control (Experiment 4)	1.06 ($p \approx 0.744$)	**1.79 ($p \approx 0.002$)**	0.71 ($p \approx 0.069$)
S-Diss-M-Faith vs. Control (Experiment 4)	**6.90 ($p < 0.001$)**	1.34 ($p \approx 0.118$)	0.90 ($p \approx 0.583$)

4.4.3.1 Results of Experiment 3

Learning Figure 4.2 shows that both the M-Harm-S-Faith and the S-Harm-M-Faith groups in Experiment 3 learned that a pattern of liquid harmony applies at their respective training distance. More specifically, the left panel of the figure compares the proportion of harmony responses given in CVLVCV-LV test items for the M-Harm-S-Faith and Control groups. The mixed logit model estimates that subjects

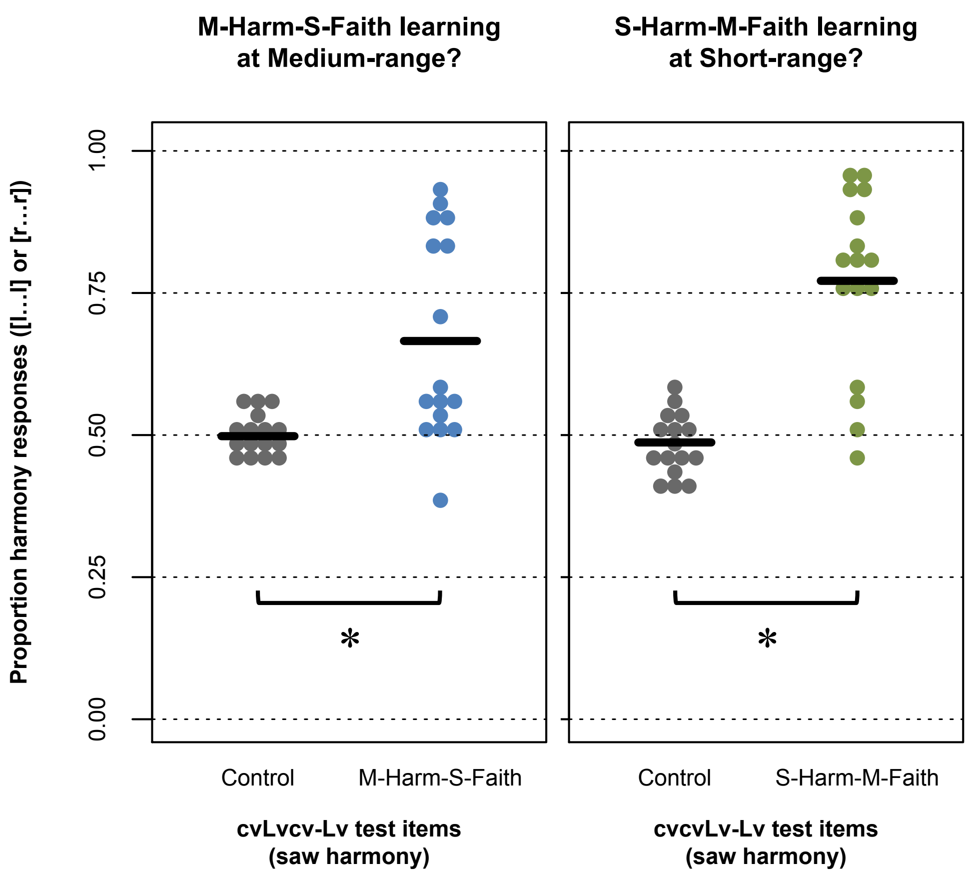

Figure 4.2: Plots comparing proportions of harmony responses for Control subjects to those of the M-Harm-S-Faith subjects in Medium-range test items (left panel) and to the S-Harm-M-Faith subjects in Short-range test items (right panel). Each dot represents individual subject performance, and group means are indicated with a horizontal line. Significance is extracted from a mixed logit model and indicates learning of the pattern each group was exposed to.

in the former group are 2.5 times (summary of all OR values can be found above in Table 4.7) more likely than a Control subject to choose harmony at Medium-range ($p < 0.001$; it is clear that this effect is driven by a subset of subjects who were more successful than the others). This result is interesting in that these subjects were exposed to evidence against harmony at the Short-range distance, but this was not enough to stop a number of participants from learning the Medium-range dependency. The right panel of Figure 4.2 shows a comparison of the S-Harm-M-Faith group and the Control group for Short-range test items. Subjects in the S-Harm-M-Faith group also picked up on the pattern of liquid harmony that was presented to them in Short-range contexts, and the model estimates that they are 7.42 times more likely than the Control group to choose harmony in CVCV$\underline{\text{L}}$V-$\underline{\text{L}}$V test items. It also appears that this group has a greater number of successful learners than the M-Harm-S-Faith group. This result is not surprising, as the pattern that the S-Harm-M-Faith group subjects were exposed to falls within the realm of TSL$_2$ languages, and we have already seen (in Experiment 1) that subjects are able to learn transvocalic liquid harmony, even without any information about the treatment of liquids at the Medium-range distance.

M-Harm-S-Faith Generalization The plots in Figure 4.3 show that the M-Harm-S-Faith group, as a whole, chose test items with harmony more often than the Control group at both the Short-range and Long-range distance. Note that the subjects in this group were not exposed to any evidence that liquid harmony should be enforced at either of these two distances. Moreover, they were given explicit evidence that liquids *do not* alternate in Short-range stems and that all possible combinations of liquids are permitted in CVCV$\underline{\text{L}}$V-$\underline{\text{L}}$V contexts. Nonetheless, the left panel of the figure shows that many subjects over-generalized the pattern, with the mixed logit model estimating that subjects in the M-Harm-S-Faith group are 2.22 times ($p < 0.001$) more likely to choose harmony than the Control group in Short-range test items. The fact that the group also tends to generalize harmony to Long-range contexts (estimated odds ratio of 1.74, $p \approx 0.011$) is further evidence that they have interpreted their training data as a pattern with unbounded locality, even though they were provided with evidence that it was not.

S-Harm-M-Faith Generalization Finally, Figure 4.4 shows results of the S-Harm-M-Faith group for both Medium-range and Long-range test items. With respect to the left panel of the figure, the mixed logit model estimates a small effect for the S-Harm-M-Faith group at Medium-range (OR = 1.57). However, this effect barely reaches statistical significance ($p \approx 0.042$), and the effect likely stems from

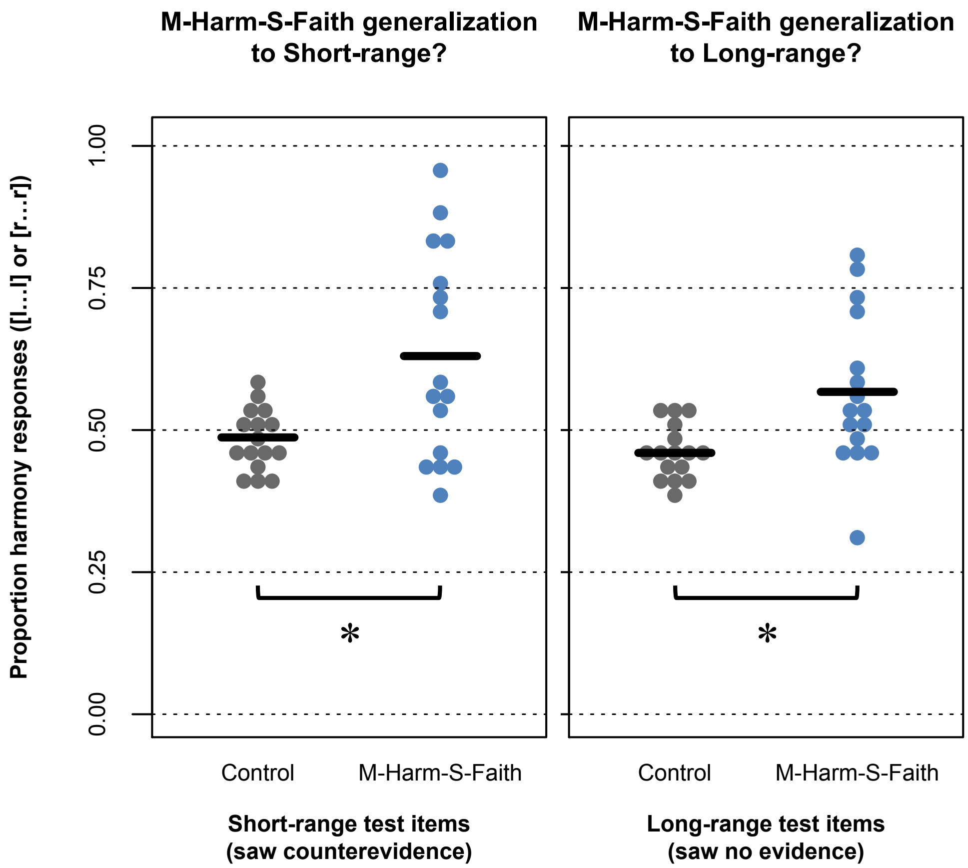

Figure 4.3: Plots comparing proportions of harmony responses for Control subjects to those of the M-Harm-S-Faith subjects in Short-range test items (left panel) and in Long-range test items (right panel). Each dot represents individual subject performance, and group means are indicated with a horizontal line. Significance is extracted from a mixed logit model and indicates generalization of the pattern the group was exposed to.

a few subjects that seem to apply the pattern of liquid harmony to Medium-range contexts. Nonetheless, this is slightly surprising in light of the fact that subjects in this group were exposed to evidence in their training that liquids in the stem stay faithful in Medium-range contexts. The right panel of the figure shows that only one subject seems to apply harmony to Long-range test items, and the effect for the group as a whole does not reach significance. The results for this group,

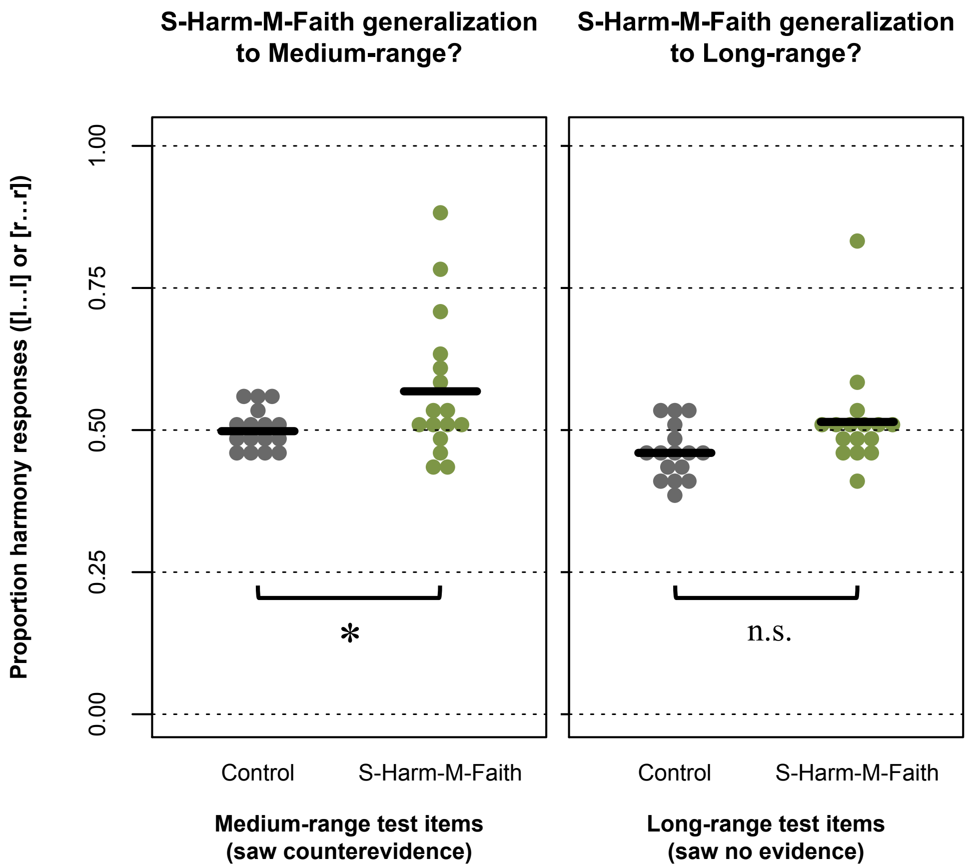

Figure 4.4: Plots comparing proportions of harmony responses for Control subjects to those of the S-Harm-M-Faith subjects in Medium-range test items (left panel) and in Long-range test items (right panel). Each dot represents individual subject performance, and group means are indicated with a horizontal line. (Non-)Significance is extracted from a mixed logit model and indicates generalization of the pattern the group was exposed to.

though peculiar when considering only whether or not an effect reaches statistical significance at the group level, are similar to the results for the S-Harm group in Experiment 1 (see Section 2.2.2.3). This suggests that subjects who are exposed to a pattern compatible with the (attested) transvocalic variant of consonant harmony are very likely to apply the pattern in only those transvocalic contexts, but that there is a small tendency (both within individuals and for the group as a whole) to apply the pattern at Medium-range as well.

4.4.3.2 Results of Experiment 4

Learning Figure 4.5 shows that both the M-Diss-S-Faith and the S-Diss-M-Faith groups in Experiment 4 learned that a pattern of liquid dissimilation applies at their respective training distance. For the M-Diss-S-Faith group, this is a relatively small effect with the mixed logit model estimating an odds ratio of about 1.79 ($p \approx 0.002$). The left panel of the figure shows that most subjects are clustered just above a 0.50 proportion of disharmony choices in Medium-range test items, and that the effect is driven by just two or three subjects who learned dissimilation with a varying degree of success. By contrast (but as expected), most subject in the S-Diss-M-Faith group, whose results are shown in the right panel of Figure 4.5, successfully learned a pattern of dissimilation for the CVCV<u>L</u>V-<u>L</u>V items that they were exposed to in training (the statistical model estimates a relatively large odds ratio of 6.90, $p < 0.001$). The overall results for the test items that represent *learning* in Experiment 4 thus do not differ from the results of Experiment 3 in terms of whether or not the effects reach statistical significance at the group level (see Figure 4.2 and the accompanying discussion). However, there does appear to be a difference in the number of individual subjects who detected the target Medium-range pattern. Out of the 16 subjects in the M-Diss-S-Faith group, only three surpass a threshold of “successful learning” (see Section 3.2.3 for a description of the binomial test used to determine whether individual subjects surpassed a 95% confidence level of having successfully learned a pattern), as compared to seven out of the 16 M-Harm-S-Faith subjects. This issue is further pursued in the discussion of individual results in Section 4.4.4 below.

M-Diss-S-Faith and S-Diss-M-Faith Do Not Generalize As shown in Figures 4.6 and 4.7, neither of the two experimental groups shows an overall tendency to generalize a pattern of liquid dissimilation to other distances. This result was expected for the S-Diss-M-Faith group, for whom the training data was compatible with a pattern of dissimilation with transvocalic locality, but differs slightly from expectations in the case of the M-Diss-S-Faith group. Specifically, given that the

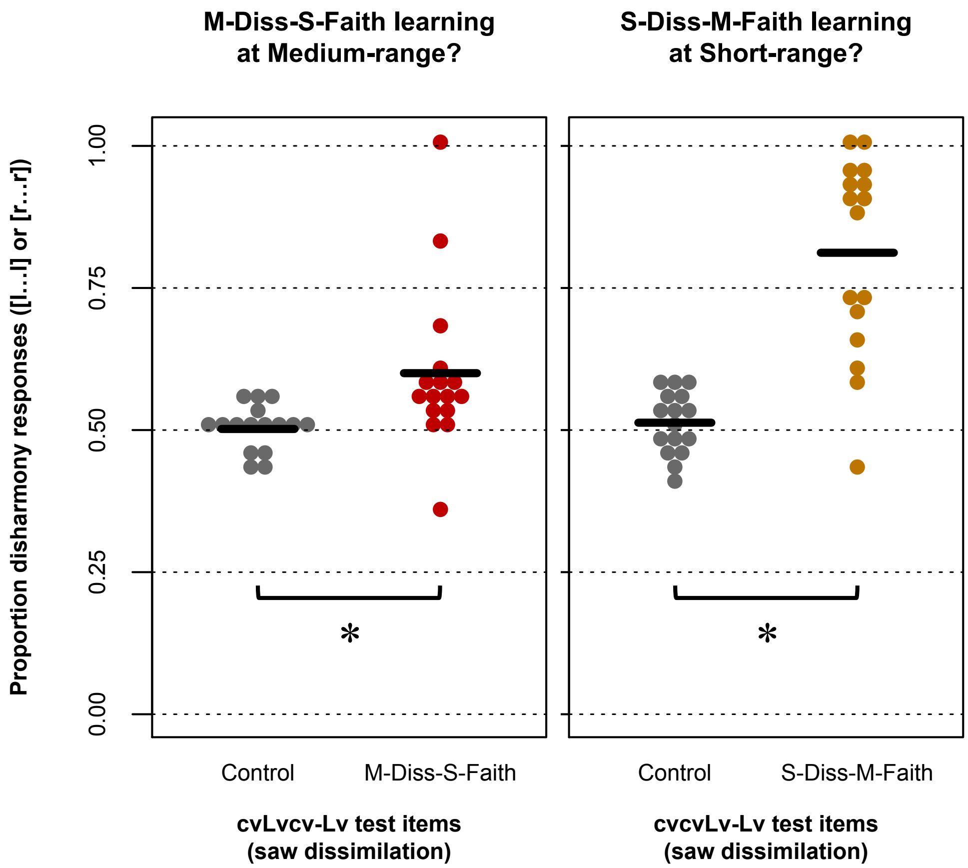

Figure 4.5: Plots comparing proportions of disharmony responses for Control subjects to those of the M-Diss-S-Faith subjects in Medium-range test items (left panel) and to the S-Diss-M-Faith subjects in Short-range test items (right panel). Each dot represents individual subject performance, and group means are indicated with a horizontal line. Significance is extracted from a mixed logit model and indicates learning of the pattern each group was exposed to.

corresponding group from Experiment 3 (M-Harm-S-Faith) showed evidence of applying harmony in an unbounded fashion, along with the fact that a few subjects in the M-Diss-S-Faith group did appear to have learned a Medium-range dependency, it is surprising that not a single subject's proportion of disharmony responses distinguishes them from the Control group in either Short- or Long-range testing items (see the individual dots in Figure 4.6). The following discussion of individ-

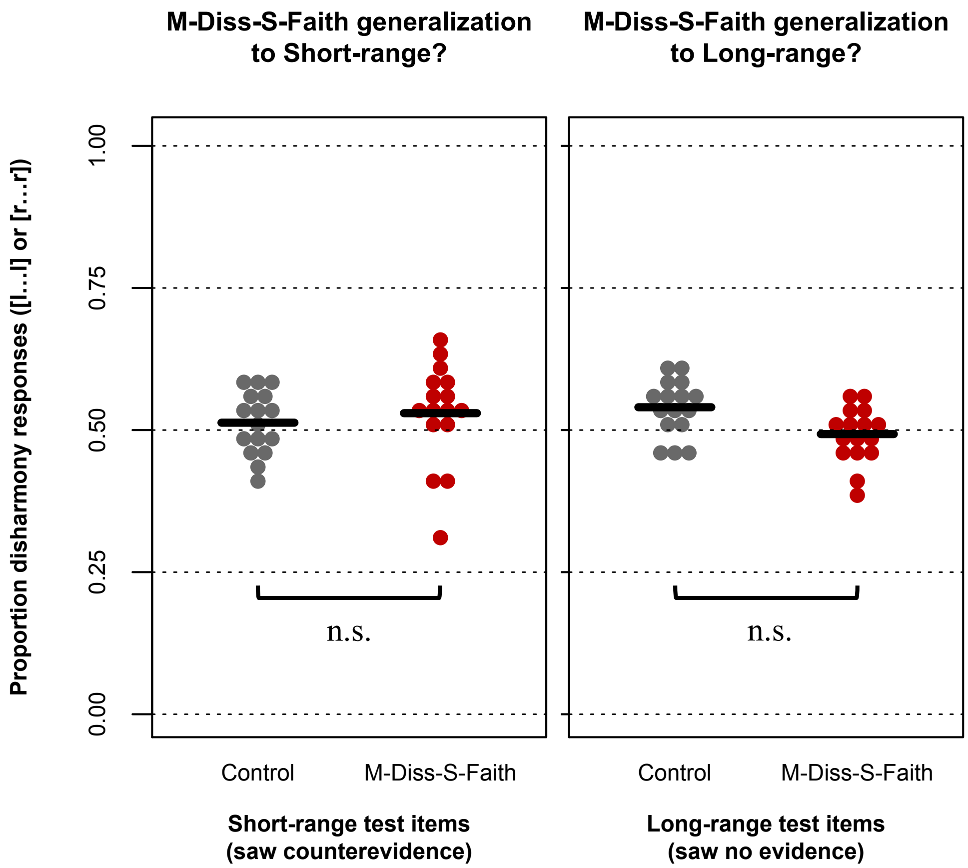

Figure 4.6: Plots comparing proportions of disharmony responses for Control subjects to those of the M-Diss-S-Faith subjects in Short-range test items (left panel) and in Long-range test items (right panel). Each dot represents individual subject performance, and group means are indicated with a horizontal line. (Non-)Significance is extracted from a mixed logit model and indicates generalization of the pattern the group was exposed to.

ual results for Experiment 4 further investigates the issue of an individual subject's ability to learn a pattern from this type of input, and presents results from an extended (yet failed) attempt to collect data for at least 12 successful learners in the M-Diss-S-Faith condition.

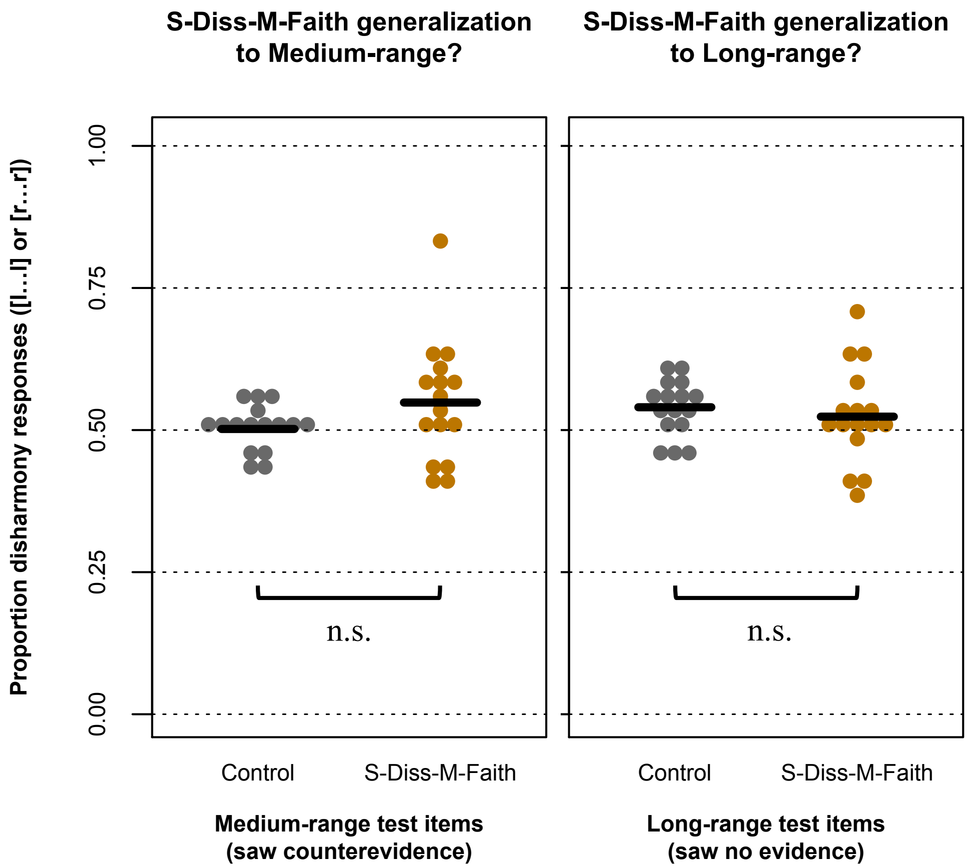

Figure 4.7: Plots comparing proportions of disharmony responses for Control subjects to those of the S-Diss-M-Faith subjects in Medium-range test items (left panel) and in Long-range test items (right panel). Each dot represents individual subject performance, and group means are indicated with a horizontal line. (Non-)Significance is extracted from a mixed logit model and indicates generalization of the pattern the group was exposed to.

4.4.4 Individual Results and General Discussion

The purpose of Experiments 3 and 4 was to investigate the possibility that adult human learners are indeed capable of learning unattested, computationally complex phonotactic patterns that are outside of the proposed hypothesis space for a human learner. As such, the present discussion focuses especially on the M-Harm-S-Faith and M-Diss-S-Faith conditions, in which the training stimuli showed evidence that liquid harmony or dissimilation holds at Medium-range distances, but not at Short-range. Note that the training data are compatible with dependencies that hold across *exactly one* intervening consonant (Medium-range only) or across *at least one* intervening consonant (Medium- and Long-range, but not Short-range), and recall that Section 4.3 demonstrated that such patterns are situated outside of the TSL_2 region. The prediction for these experiments was therefore that the M-Harm-S-Faith and M-Diss-S-Faith groups would either not learn any dependency between liquids, or that they would erroneously learn an unbounded pattern that can be characterized in TSL_2 terms. In the overall group results of Experiment 3, the latter was seen for the subjects in the M-Harm-S-Faith group, who were statistically more likely than the Control group to choose harmony no matter whether the liquids were in a Short-, Medium-, or Long-range test item. Of further interest are the individual subject results for the M-Harm-S-Faith group (illustrated in Figure 4.8). Out of the seven subjects with the highest proportions of harmony responses at the Medium-range distance, which is where harmony occurred in their training data, six of them appear to have generalized the pattern into Short-range contexts in spite of the counter-evidence provided. Only a few of these subjects also generalized to Long-range test items, though this result is similar to the findings of Experiments 1 and 2 in that subjects seem reluctant to extend the alternation to consonants in word-initial position.

Recall that of the 16 subjects in the M-Diss-S-Faith group in Experiment 4, only three surpassed the defined threshold of learning a dependency. Of these, no subject generalized the pattern of dissimilation to either the Short-range or the Long-range test items. In order to boost the number of learners in this group to get a more reliable picture of subject behaviour, more data was collected in an attempt to identify a minimum of 12 successful learners in the M-Diss-S-Faith group (as was done for the alternative analyses for Experiments 1 and 2; see Section 3.2.3). However, after running a total of 40 native English speakers in this training condition, the number of learners only rose to eight. As seen in Figure 4.9, nearly all of the subjects cluster around a 0.50 proportion of choosing dissimilation at all three testing distances. Of those eight who do successfully learn the Medium-range dissimilation, five surpass the same threshold at the Short-range distance: subjects 736, 766, 757, 752,

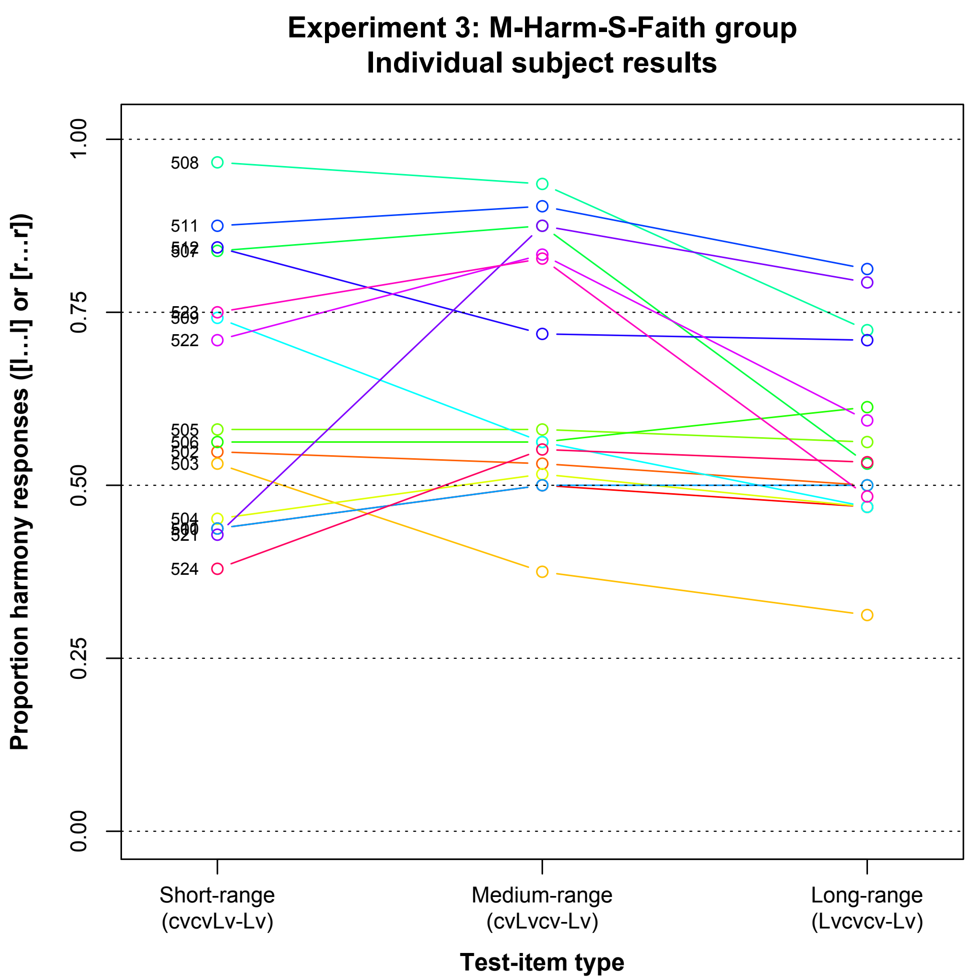

Figure 4.8: Individual results for M-Harm-S-Faith group at each of the three testing distances. Individual subjects are distinguished by colour and 3-digit code.

and 732. In this sense, the additional data that was collected beyond the original 16 subjects offers much better support for the prediction that subjects who were exposed to a beyond-transvocalic pattern should over-generalize into the Short-range contexts, as was also seen in the results of the M-Harm-S-Faith group.

With respect to the range of patterns observed for individual subjects in Experiments 3 and 4, several subjects exhibit certain unexpected behaviours. For example,

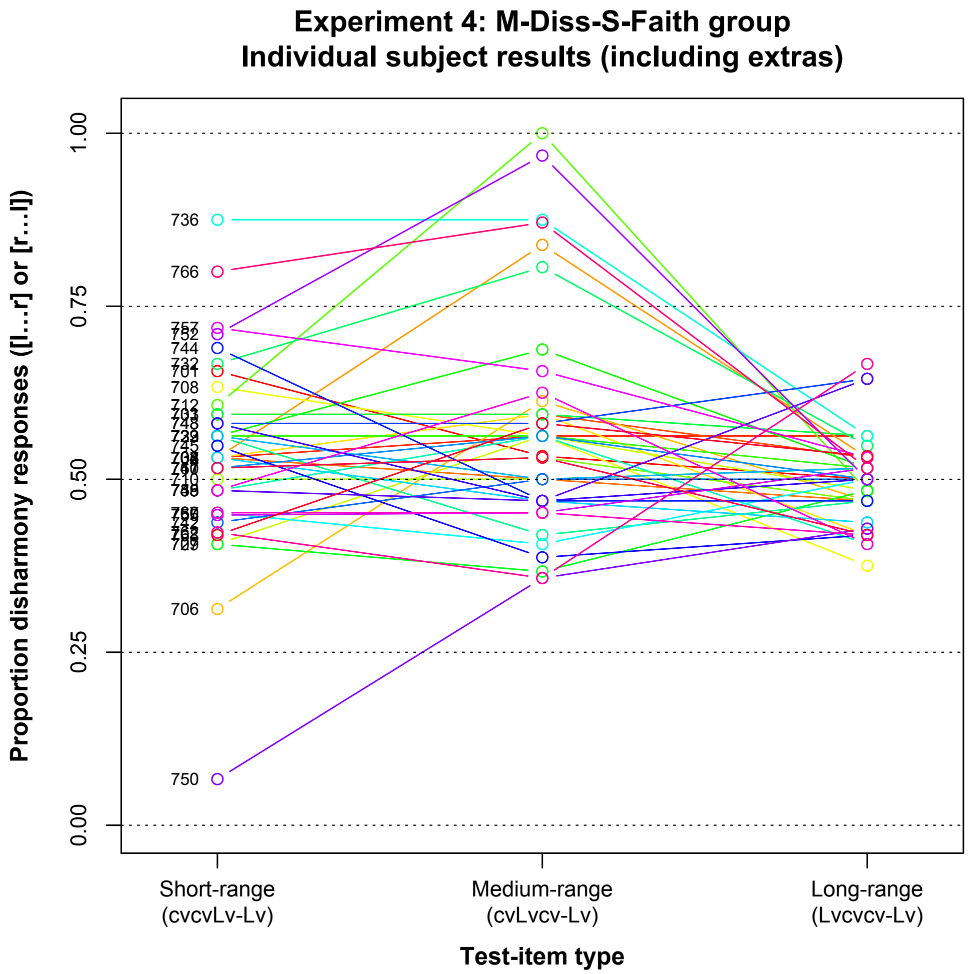

Figure 4.9: Individual results for M-Diss-S-Faith group at each of the three testing distances. Individual subjects are distinguished by colour and 3-digit code.

subject 521 of the M-Harm-S-Faith group (results in light purple in Figure 4.8) appears to have learned a beyond-transvocalic version of liquid harmony—a pattern that is not predicted to be possible, even within the ABC framework. Subject 750 (in Figure 4.8, also in light purple) seems to have learned a pattern of transvocalic liquid *harmony*, despite having been exposed to *dissimilation* as a member of the M-Diss-S-Faith training group. Such peculiarities bring up a number of important points. First and foremost is that results like these are not statistically reliable, and the experiments were not designed to draw post hoc conclusions from individual subjects that deviate from overall group behaviour. Second, we do not yet know enough about what subjects are actually doing in artificial language learning tasks to assess subjects on an individual basis. It may be, for example, that certain subjects access a completely different learning module for a superficial language learning task. Nonetheless, I think it is safe to assume that some humans learn languages differently than others, and furthermore that some humans may be capable of acquiring patterns that are more complex than those typically found in natural language. This idea calls into question what it means to be a learning bias, which up to this point, I have discussed as a boundary that applies for all language learners. While I do not pursue the issue any further in this dissertation, it may be that learning biases are better conceived of in probabilistic terms, such that complex phonotactic patterns (should they arise in the first place) could be acquired as-is by certain learners. However, an overall tendency for the average learner to misinterpret their input as evidence for a simpler pattern (e.g. learning an unbounded pattern from beyond-transvocalic input) makes it highly unlikely that such a pattern could persist for even one generation. From this view, it may be better to think of the TSL_2 region not as a strict learning bias per se, but as a region of patterns that are likely to be learned by everyone, and therefore be stable over time.

4.5 Summary and Conclusions

In pursuing an answer to the question of what constitutes a possible, human-learnable phonotactic pattern, it is important to establish a well-defined boundary that divides patterns into those that are predicted to be in the human learner's hypothesis space and those that are not. The central claim of this chapter (and indeed of this dissertation more generally) is that the class of Tier-based Strictly 2-Local formal languages (as outlined in Heinz et al., 2011) offers an excellent approximation of such a boundary. I have supported this argument in a number of ways.

From a typological perspective, the cross-linguistic properties of locality and blocking can be captured in TSL_2 terms, as the difference between unbounded patterns, transvocalic patterns, and patterns with blocking is easily generated by varying the contents of the tier specified by the formal grammar without the need for modification to any other parameters. Furthermore, there are many unattested patterns (including several pathologies predicted by the Agreement by Correspondence framework) whose formal properties demonstrably situate them outside of the TSL_2 region. With respect to experimental evidence, the overall results of Experiments 1 through 4 indicate that subjects are indeed able to learn patterns that are compatible with a TSL_2 language in terms of locality. However, when exposed to more complex patterns (as in the case of the M-Harm-S-Faith and M-Diss-S-Faith training conditions), very few subjects are able to detect any pattern whatsoever, and those that do often show evidence of having learned a TSL_2 grammar that contradicts their training data.

When considered as a whole, I take the above findings as support for the proposal that the human phonotactic learner is equipped with an analytic learning bias that restricts the hypothesis space to patterns that fall within the TSL_2 region (or at least something very close to it). Furthermore, the evidence suggests that this bias manifests itself in the form of typological gaps consisting of patterns that are logically possible, but that remain both synchronically and diachronically inaccessible because they cannot be learned as TSL_2 patterns.

Chapter 5

Questions About the TSL_2 Approach

The previous chapter argued that the region of Tier-based Strictly 2-Local formal languages is a reasonable, if not excellent formal definition of the boundary that establishes a set of possible, human-learnable phonotactic patterns and separates them from complex patterns that are not human-learnable (and therefore impossible). This chapter scrutinizes the TSL_2 proposal by asking three questions:

1. Is the TSL_2 region too big?
2. Is the TSL_2 region computationally learnable?
3. Is the TSL_2 region too small?

Sections 5.1, 5.2, and 5.3 investigate each of the above questions in turn, taking into account evidence from typological surveys of long-distance consonant interactions, the results of previous artificial language learning studies from the literature, and investigations of the computational properties of the TSL_2 class of formal languages. Section 5.4 concludes that characterizing phonotactic patterns as TSL_2 stringsets remains a viable strategy in accounting for a wide variety of empirical data.

5.1 Is the TSL_2 Region Too Big?

A potential reason to be skeptical of the TSL_2 approach is that, formally speaking, a tier T can be any combination of segments in Σ despite the fact that most attested long-distance phonotactic dependencies seem to hold on a tier that could be defined by a phonological feature or natural class, rather than an arbitrary set of segments. For example, in each of the grammars for the three variants of sibilant harmony presented in Table 5.1 (adapted from above), the members of T can be described as a natural class, shown in parentheses.

Table 5.1: TSL_2 grammars for three attested variants of sibilant harmony

Type of pattern	Σ	T	R
Unbounded	{s, ʃ, p, t, a}	{s, ʃ} (sibilants)	{*sʃ, *ʃs}
Blocking	{s, ʃ, p, t, a}	{s, ʃ, t} (coronal obstruents)	{*sʃ, *ʃs}
Transvocalic	{s, ʃ, p, t, a}	{s, ʃ, p, t} (consonants)	{*sʃ, *ʃs}

I reiterate that this is a significant deviation from the concept of a tier (or projection of segments), which has long been used in theoretical frameworks such as feature geometry or autosegmental theory (e.g. Clements, 1980, 1985; Shaw, 1991; Odden, 1994; Blevins, 2004; Clements and Hume, 1995), which have seen relative success in accounting for the range of attested patterns in terms of the classes of interacting segments. It may therefore be desirable to integrate certain aspects of phonological theory into the TSL_2 account of long-distance phonotactics in order to further limit the range of possible patterns. However, determining the best way to do so remains an open problem for formal-language-theoretic approaches to phonology more generally, and is outside the scope of the present research. Rather than pursue such an addition to the theory in this dissertation, I instead argue that the relative flexibility of tier specification is a potential *advantage* of the TSL_2 approach because it offers a simple account of peripheral empirical data. Specifically, this section presents patterns from two languages, Kinyarwanda (Section 5.1.1) and Latin (Section 5.1.2), each of which contains a long-distance phonotactic dependency that belongs to the TSL_2 class, but whose grammar necessarily specifies *T* with a set of segments that cannot be described as a natural class. I then summarize the experiments of Koo and Oh (2013), who provide evidence that adult human learners are able to detect arbitrarily defined non-adjacent dependencies in C̲vcvC̲ contexts (Section 5.1.3).

5.1.1 Kinyarwanda Sibilant Harmony With Blocking

As an example of a TSL_2 pattern of agreement whose tier cannot be characterized as a natural class of segments, I present data from a sibilant harmony pattern found in Kinyarwanda, which is blocked by certain intervening segments (see Walker and Mpiranya, 2005; Walker et al., 2008; Hansson, 2007, 2010a). Unless otherwise noted, all cited data are from Walker and Mpiranya (2005). As shown in (1), regressive retroflexion harmony among sibilant fricatives is obligatory in …S̲v(:)S̲… contexts, such that no segment belonging to the set [s, z, n͡z] may precede any of the retroflex [ʂ, ʐ, ɳ͡ʐ] in a transvocalic configuration. Note that the retroflex trigger may be derived from a following /i/, as illustrated below with the agentive suffix /-i/ and the perfective suffix, which is represented by /-i-e/ (following Walker and Mpiranya, 2005). As evidence that harmony is purely anticipatory and only triggered by the series of retroflex sibilant fricatives, Walker and Mpiranya (2005) provide the form /-ʂit-i-e/, which surfaces as [-ʂise] ‘penetrated (perf.)’, rather than *[-sise] or *[-ʂiʂe].

(1) Kinyarwanda: obligatory …S̲v(v)S̲… harmony

a.	/-sas-i/	-**ʂ**a**ʂ**i	‘bed maker’	*-**s**a**ʂ**i
b.	/-so:n͡z-i/	-**ʂ**o:**ɳ͡ʐ**i	‘victim of famine’	*-**s**o:**ɳ͡ʐ**i
c.	/-úzuz-i-e/	-ú**ʐ**u**ʐ**e	‘filled (perf.)’	*-ú**z**u**ʐ**e
d.	/-sá:z-i-e/	-**ʂ**á:**ʐ**e	‘became old (perf.)’	*-**s**áa**ʐ**e

Harmony is also optional beyond the transvocalic window, as seen in (2) where harmony is permitted (though not obligatory) across intervening non-coronal consonants such as [k, g, m].

(2) Kinyarwanda: optional …$\underline{\text{S}}$…c…$\underline{\text{S}}$… harmony

a.	/-sákuz-i-e/	-**ʂ**áku**ʐ**e (*also* -**s**áku**ʐ**e)	‘shouted (perf.)’
b.	/-zímagiz-i-e/	-**ʐ**ímagi**ʐ**e (*also* -**z**ímagi**ʐ**e)	‘misled (perf.)’
c.	/-ásamuz-i-e/	-á**ʂ**amu**ʐ**e (*also* -á**s**amu**ʐ**e)	‘made open mouth (perf.)’

For expository purposes, the present discussion treats the above pattern as a categorical case of unbounded sibilant harmony. I note that optionality is not itself problematic in the TSL_2 approach, since we can simply label both options as phonotactically grammatical strings. However, in cases where harmony is obligatory at one level of locality, as is the case for the Kinyarwanda data in (1), but not at another, as in (2), the distinction cannot be captured as a single TSL_2 language. I consider issues relating to patterns with multiple tiers in Section 5.3, but the treatment of optionality and gradience within the TSL_2 framework is left for future research.

Considering the basic generalizations in (1) and (2), it is easy to see that in a TSL_2 grammar (which requires harmony at all distances), the set of tier-based 2-factor restrictions would be R = {*sʂ, *sʐ, *sɳ͡ʐ, *zʂ, *zʐ, *zɳ͡ʐ, *n͡zʂ, *n͡zʐ, *n͡zɳ͡ʐ}. However, the grammar cannot simply specify the tier as consisting of the sibilant fricatives, with T = {s, z, n͡z, ʂ, ʐ, ɳ͡ʐ}, as the data in (3)-(5) demonstrate that harmony is blocked by a number of other consonants (unfaithful alternations resulting in harmony are not even optionally permitted when a blocker intervenes between the trigger and target).

(3) Kinyarwanda: harmony blocked by non-retroflex coronal stops/nasals

a.	/-síːtaːz-i-e/	-**s**íː**t**aː**ʐ**e (*-**ʂ**íː**t**aː**ʐ**e)	‘made stub (perf.)’
b.	/-sódoːk-i-izʐe/	-**s**ó**d**oːke**ʐ**e (*-**ʂ**ó**d**oːke**ʐ**e)	‘made move slowly (perf.)’
c.	/-súnuːk-i-izʐe/	-**s**ú**n**uːki**ʐ**e (*-**ʂ**ú**n**uːki**ʐ**e)	‘showed furtively (perf.)’

(4) Kinyarwanda: harmony blocked by palatals (note: /n+i/ → [ɲ])

a. /-zúja:z-i-e/ -**z**ú**j**a:**ʐ**e ‘became warm (liquid) (perf.)’
(*-**ʐ**ú**j**a:**ʐ**e)

b. /-zíg-an-i-izʐe/ -**z**íga**ɲ**i**ʐ**e ‘economized (perf.)’
(*-**ʐ**íga**ɲ**i**ʐ**e)

(5) Kinyarwanda: [t͡s] is not a target and blocks harmony

a. /-t͡si:mbaɽaz-i-e/ -**t͡s**i:mbaɽa**ʐ**e ‘made obstinate (perf.)’
(*-ʈ͡ʂi:mbaɽa**ʐ**e)

b. /-set͡saguz-i-e/ -**s**e**t͡s**agu**ʐ**e ‘made carve up (perf.)’
(*-**ʂ**e**t͡s**agu**ʐ**e, *-**ʂ**e**ʈ͡ʂ**agu**ʐ**e)

The data in (5) are especially interesting. (5a) shows that the non-retroflex coronal affricate [t͡s] is not targeted by harmony despite being contrastive with a retroflex counterpart [ʈ͡ʂ], and (5b) shows that [t͡s] blocks harmony when it intervenes between two sibilant fricatives. The importance of this is that it eliminates any possibility of defining the set of blockers as (non-retroflex) coronals that are not contrastive for retroflexion. At present then, since we know that blockers must also be members of the tier in a TSL_2 grammar, we know that *T* is comprised of (at least) the following segments: {s, z, n͡z, ʂ, ʐ, ɳ͡ʐ, t, d, n, j, ɲ, t͡s}. The final data that we need to consider before inducing the correct TSL_2 grammar are presented in (6) and (7). Note that (7) is cited from Walker et al. (2008).

(6) Kinyarwanda: [ɽ] is transparent to harmony, and is not a trigger

a. /-togoseɽez-i-e/ -togo**ʂ**eɽe**ʐ**e ‘made boil for (perf.)’
(*also* -togo**s**eɽe**ʐ**e)

b. /-seɽuz-i-e/ -**ʂ**eɽu**ʐ**e ‘provoked, irritated (perf.)’
(*also* -**s**eɽu**ʐ**e)

c. /-soɽ-a/ -soɽa ‘pay tax’
(*-ʂoɽa)

d. /-ziɽ-a/ -ziɽa ‘be forbidden (taboo)’
(*-ʐiɽa)

(7) Kinyarwanda: harmony is blocked by [ɳ͡ɖ]

[βa**s**a:**ɳ͡ɖ**a**ʐ**e] ‘they blew up’ *[βa**ʂ**a:**ɳ͡ɖ**a**ʐ**e]

In the above data, (6a) and (6b) show that the sonorant [ɽ] is transparent to the (optional) retroflexion harmony among sibilant fricatives, suggesting that [ɽ] is not in *T*. Further evidence for this is that sequences of …sVɽ… and …zVɽ… are permitted in (6c) and (6d). As such, [ɽ] is not a trigger of retroflexion harmony, and does not need to be included in any tier-based 2-factors in *R*. By contrast, the prenasalized stop [ɳ͡ɖ], which Walker et al. (2008) found to be phonetically retroflex (cf. the transcription /n͡d/ in Walker and Mpiranya, 2005), does block retroflexion harmony, as seen above in (7), and must be included in *T*. The final TSL_2 grammar for Kinyarwanda is given in (8).

(8) $$G = \left\langle T = \left\{ \begin{array}{c} \text{s, z, n͡z, ʂ, ʐ, ɳ͡ʐ} \\ \text{t, d, n, j, ɲ, t͡s, ɳ͡ɖ} \end{array} \right\}, R = \left\{ \begin{array}{c} \text{*sʂ, *sʐ, *sɳ͡ʐ, *zʂ, *zʐ} \\ \text{*zɳ͡ʐ, *n͡zʂ, *n͡zʐ, *n͡zɳ͡ʐ} \end{array} \right\} \right\rangle$$

Since the segment [ɽ] is not a member of the tier specified by the grammar in (8), there is no natural class that can be used to define *T*. Any attempt to do so will either erroneously include [ɽ] despite its demonstrable transparency (for example, *T* cannot be the coronal consonants), or exclude consonants that we know to be in *T* (for example, *T* cannot be the coronal obstruents or the coronal non-continuants). In order to approximate *T* with a natural class, we would have to stipulate that something like 'the non-rhotic coronal consonants' is a natural class of segments. While I know of no version of distinctive feature theory in which this holds true, it would be precisely a pattern such as this that might motivate such a proposal. However, if arbitrary collections of segments are not a problem for the learner, then we may not need to force a class of non-rhotic coronal consonants into a theory of natural classes for the sole purpose of accounting for rare patterns. Finally, I point out alternative ways of using natural classes to describe the contents of *T* for Kinyarwanda. We could, for example, either reference a union of several natural classes (such as $T = \{\text{sibilants}\} \cup \{\text{coronal stops}\} \cup \{\text{palatal consonants}\}$), or use some other operation over sets of segments that are natural classes, such as $T = \{\text{coronal consonants}\} - \{\text{ɽ}\}$ (using any preferred combination of features that picks out [ɽ] as a natural class consisting of one segment).

5.1.2 Latin Liquid Dissimilation

As another example of a long-distance dependency that can be characterized with a TSL_2 grammar but whose tier cannot be described as a natural class of segments, this section presents data from the well-known case of the Latin -*alis* ~ -*aris* alternation. While there has been a historical lack of consensus in the literature with respect to the precise details of the pattern (see, e.g., Watkins, 1970; Dressler, 1971; Jensen, 1974; Steriade, 1987a), I follow a more recent description of the synchronic aspects of this allomorphy given by Cser (2010; all data below are cited from this study unless otherwise noted). Cser performs a corpus study of Latin focusing on the phonotactics in Classical and Post-classical Latin, between the 1st century BC and the 4th century AD.

The commonly cited, analyzed, and counter-exemplified generalization of the pattern found in Latin is that underlying /-al/ surfaces as [-ar] (e.g. the masc.nom.sg. form [-ar-is]) when preceded by an [l], unless [r] intervenes. The data in (9) provide evidence that the underlying form of the suffix is /-al/, which surfaces faithfully when there is no [l] in the stem. The basic dissimilation is seen in (10), when the stems contain a liquid [l] with no [r] following it. The data in (11) show that dissimilation is blocked when the stem contains an *l*…*r* subsequence.

(9) Latin: default form of suffix is -*alis*

- a. *nav-**al**-is* ‘naval’
- b. *autumn-**al**-is* ‘autumn-’
- c. *hiem-**al**-is* ‘winter-’
- d. *reg-**al**-is* ‘royal’

(10) Latin: dissimilation triggered by preceding [l]

- a. *consu**l**-**ar**-is* ‘consular’
- b. *popu**l**-**ar**-is* ‘popular’
- c. *ste**ll**-**ar**-is* ‘stellar’
- d. *mi**l**it-**ar**-is* ‘military’
- e. ***l**un-**ar**-is* ‘lunar’

(11) Latin: dissimilation blocked by intervening [r]

a. *f**l**o**r**-a**l**-is* ‘floral’

b. *p**l**u**r**-a**l**-is* ‘plural’

c. ***l**ate**r**-a**l**-is* ‘side-, lateral’

The above description of the pattern is characteristic of those given in much of the literature (e.g. Watkins, 1970; Dressler, 1971; Jensen, 1974; Steriade, 1987a), with some authors making note of several counterexamples (such as *legalis*, cf. **legaris*) in which dissimilation does not occur when expected. With respect to these basic generalizations for (9)-(11), the pattern is easily described as TSL_2, with $G = \langle T = \{\mathrm{l}, \mathrm{r}\}, R = \{\text{*ll}\}\rangle$ (closely resembling the pattern in Georgian; see Section 4.1.2 above). However, Cser's (2010) corpus investigation reveals that these ‘exceptions’ simply follow an additional component of the pattern: dissimilation is also blocked by non-coronal consonants. This extension includes both labial and velar consonants, as shown in (12) with an example for each of [b, m, w, k, g]. (Note that all examples in this section use an orthographical representation of Latin, in which ‘*v*’ and ‘*c*’ correspond to [w] and [k], respectively.)

(12) Latin: dissimilation blocked by intervening non-coronal consonants

a. *g**l**e**b**-a**l**-is* ‘consisting of clods’

b. *fu**lm**in-a**l**-is* ‘projectile’

c. *p**l**u**v**i-a**l**-is* ‘rainy’

d. *umbi**l**i**c**-a**l**-is* ‘umbilical’

e. ***l**e**g**-a**l**-is* ‘legal’

Aggregating all of the above data, we have evidence that each of [l, r, b, m, w, k, g] must be on the tier, and that [t, n] cannot be on the tier, since (10d) and (10e) demonstrate that they are transparent (in addition to the vowels being transparent). With respect to additional consonants such as [s, d, h, p, f], Cser (2010) cites no examples in which they occur in the necessary context. However, even in a best case scenario in which homorganic consonants behave in the same way, the TSL_2 grammar for the pattern found in Latin would be as shown in (13).

(13) $G = \left\langle T = \left\{ \begin{array}{c} \text{l, r, p, b, f, m,} \\ \text{w, k, g, (h)} \end{array} \right\}, R = \{\text{*ll}\} \right\rangle$

The above tier, which includes all labial and velar consonants but only two coronals, clearly cannot be described as a natural class of segments. I therefore take the pattern found in Latin as further evidence that such grammars should not be absent from the human learner's hypothesis space.

5.1.3 Experimental Learning of Arbitrary Tiers

As a different type of empirical support for retaining arbitrary patterns in the learner's hypothesis space, I summarize the experimental findings of Koo and Oh (2013), who argue that human learners are capable of detecting phonotactic regularities among sets of segments that cannot be phonologically defined (e.g. using features or natural classes).

Koo and Oh (2013), a methodologically improved replication of Koo and Callahan (2012), present the results of two artificial language learning studies in which native Korean-speaking subjects were exposed to long-distance dependencies between consonants C_1 and C_3 in words of the form $C_1VC_2VC_3$. For each of the words, C_1 was one of three possible consonants {m, tʰ, tʃ}, C_2 was one of {h, k, s}, and C_3 was one of {ŋ, l, p}. In their Experiment A, words of the form [**m**VCV**l**], [**tʰ**VCV**ŋ**], and [**tʃ**VCV**p**] were legal, and all others were not (illegal words were absent in the training phase). In their Experiment B, the dependency still held between C_1 and C_3, but subjects were instead exposed to a different permutation of grammatical C_1, C_3 pairs, consisting of words of the form [**m**VCV**ŋ**], [**tʰ**VCV**p**], and [**tʃ**VCV**l**].

Both sets of subjects completed a testing phase in which they were asked to rate the familiarity of a novel stimulus on a scale of 1 (least familiar) to 5 (most familiar). The interesting aspect of the study is that both sets of subjects completed the exact same testing phase, such that 50% of the novel test items adhered to the phonotactic dependency of Experiment A (and were therefore illegal words in Experiment B), and 50% followed the pattern in the training phase of Experiment B (and were therefore illegal words for Experiment A). Results of their study indicate that subjects in both experiments learned the respective patterns (rating the "legal" words as more familiar than the “illegal” words), which they take as evidence against the idea that human learners can only learn dependencies among segment pairs that are adjacent on tiers defined by natural classes.

To translate Koo and Oh’s results into TSL_2 terms, subjects in Experiment A learned a pattern that can be represented by the grammar in (14), while subjects in Experiment B learned a pattern that is generated by the grammar in (15).

(14) $$G_A = \left\langle T_A = \left\{ \begin{array}{c} \text{m, t}^\text{h}\text{, tʃ,} \\ \text{ŋ, l, p} \end{array} \right\}, S_A = \{\text{ml, t}^\text{h}\text{ŋ, tʃŋ}\} \right\rangle$$

(15) $$G_B = \left\langle T_B = \left\{ \begin{array}{c} \text{m, t}^\text{h}\text{, tʃ,} \\ \text{ŋ, l, p} \end{array} \right\}, S_B = \{\text{mŋ, t}^\text{h}\text{p, tʃl}\} \right\rangle$$

Since there were no restrictions on the distribution of {h, k, s} in the C_2 position for either experiment, each of those three segments must be transparent, and cannot be a member of the tier specified by the above grammars. This means that both of the learned tiers ($T_A = T_B$) must include all of {m, tʰ, tʃ, ŋ, l, p} but exclude {h, k, s} (as well as the vowels)—a distinction that cannot be made using natural classes. Finally, note that not only are the tiers arbitrarily defined in (14) and (15), but the sets of tier-based 2-factors that are permitted (or restricted) are likewise arbitrary. For most attested phonotactic patterns, the sets *S* or *R* are easily described in terms of phonological theory, including many of the patterns presented as examples throughout this dissertation (e.g. avoid 2-factors with the features [+ant][–ant], or only permit 2-factors that have different places of articulation, etc). Though there do not seem to be many (or any) natural languages that contain such an extreme example of a dependency with an arbitrarily defined set of 2-factors in *S* or *R*, the results of Koo and Oh (2013) nonetheless indicate that subjects are capable of detecting such patterns in an artificial language.

5.1.4 Conclusion: The TSL_2 Region is Not Too Big

Before concluding that the class of TSL_2 languages does not include too many patterns, it is worth pointing out that the patterns found in Kinyarwanda (Section 5.1.1) and Latin (Section 5.1.2), as they are described above, remained undetected for quite some time even though the data was already accessible. In the case of Kinyarwanda, recall that sibilant harmony is obligatory in transvocalic contexts but optional outside of that window. While both Kimenyi (1979) and Coupez (1980) describe this basic generalization, the optional application of the pattern obscured the fact that certain segments categorically block harmony and it was not until later work that a complete description of the pattern was provided (Walker and Mpiranya, 2005; Walker et al., 2008). With respect to the pattern found in Latin, words such as *legalis* and *umbilicalis* were long cited as exceptions to an otherwise intuitive description of the data, and (to my knowledge) it was not until Cser's (2010) corpus study that a further aspect of the pattern—that labial and velar consonants block the dependency—was detected in the data. It is therefore conceivable, and perhaps likely, that there exist many examples of relatively arbitrary phonotactic dependencies that have not yet been discovered, precisely because of their unexpected nature. Furthermore, the proposed learning bias that restricts the range of possible languages to the TSL_2 region need not be the only bias at play, and there may be other reasons for the cross-linguistic underrepresentation of arbitrary patterns. That is, even if the overwhelming majority of phonotactic patterns can be described as a restriction against a phonologically defined set of 2-factors on a tier that is a natural class, this is not necessarily reason to omit others from a model of the learner's hypothesis space, and a resolution of the issue depends on further empirical data (e.g. results from experimental studies of human learning). Importantly, the set of patterns predicted when allowing arbitrarily defined tiers is a superset of those predicted when imposing some restriction on the potential contents of a tier. While it would be trivial to reduce the learner's hypothesis space in accordance with some proposed set of restrictions on what can be a tier, I do not pursue this option since, as the following section demonstrates (following Jardine and Heinz, 2015), it is not necessary to do so for reasons of computational learnability.

5.2 Is the TSL_2 Region Computationally Learnable?

I have now argued that the class of TSL_2 languages offers a close approximation of the types of patterns found in natural language, and the patterns that humans are (not) able to learn in the laboratory. I have also shown that the relative complexity of the grammar, which requires specification of both a tier and a set of restrictions against certain 2-factors, is necessary in order to account for attested properties of locality and blocking. Furthermore, Experiments 3 and 4 (see Section 4.4) offered support for leaving certain dependencies that do not belong to the TSL_2 language class outside of the human learner's hypothesis space (i.e. those that apply only beyond the transvocalic window), and the results of Koo and Oh (2013) were cited as evidence in favour of leaving phonologically arbitrary TSL_2 patterns within the range of human learners. However, the number of possible TSL_2 grammars is enormous for a segment inventory of the typical size for a natural language. This presents a significant challenge to the proposal that the entire TSL_2 region could be traversed efficiently by a human language learner in order to arrive at the correct grammar for any such pattern. After expanding on the precise nature of the TSL_2 learning problem, Section 5.2.1 summarizes a formal learning algorithm recently proposed by Jardine and Heinz (2015) that can provably and efficiently learn the class of TSL_2 languages. Section 5.2.2 then provides an example stepwise implementation of their algorithm, as applied to a schematic case of sibilant harmony with blocking modelled after Slovenian (Jurgec, 2011; see Section 4.2.1).

To better illustrate the problems posed to a potential learner of a TSL_2 language, I present a calculation of the number of possible grammars for the simplified segment inventory used in Table 4.2: $\Sigma = \{\text{s}, \text{ʃ}, \text{p}, \text{t}, \text{a}\}$. Since there are five segments in Σ, there are exactly 32 combinations of segments in Σ that could constitute the tier *T*. More generally, for some Σ there are $2^{|\Sigma|}$ possible tiers, where $|\Sigma|$ is the number of segments in the inventory. Note that of the 32 options for *T* in this example, one includes all five segments ($T_1 = \{\text{s}, \text{ʃ}, \text{p}, \text{t}, \text{a}\}$), five include four segments ($T_2 = \{\text{s}, \text{ʃ}, \text{p}, \text{t}\}$, $T_3 = \{\text{s}, \text{ʃ}, \text{p}, \text{a}\}$, $T_4 = \{\text{s}, \text{ʃ}, \text{t}, \text{a}\}$, $T_5 = \{\text{s}, \text{p}, \text{t}, \text{a}\}$, $T_6 = \{\text{ʃ}, \text{p}, \text{t}, \text{a}\}$), ten include three segments, ten include two segments, five include one segment, and the remaining logical possibility is $T_{32} = \emptyset$ (where $\emptyset$ denotes the empty set, with $|T_{32}| = |\emptyset| = 0$). However, since a TSL_2 grammar is a two-tuple that includes specification of both *T* and *R*, each of the 32 possible tiers also has a number of possibilities for the accompanying set of 2-factor restrictions in *R*. For $T_1 = \{\text{s}, \text{ʃ}, \text{p}, \text{t}, \text{a}\}$, there are $|T_1|^2 = 25$ possible 2-factors (i.e. {ss, sʃ, sp, st, sa, ʃs, ʃʃ, … , ap, at, aa}), each of which may or may not be included in *R*, resulting in $2^{25} = 33\ 554\ 432$ possible grammars when $T = T_1$. Likewise, when $|T| = 4$, there are 2^{16} possible sets *R*, and so on (where the number of possible specifications of *R*, given some *T* is

$2^{|T|^2}$). I complete the present example by adding up the number of possible grammars for each of T_1 through T_{32}, as shown in Table 5.2. The result, that the number of possible TSL$_2$ grammars reaches nearly 34 million with an inventory of just five segments, raises the concern that it may be impossible for a learner to traverse the space of TSL$_2$ languages efficiently in order to arrive at the correct grammar.[1]

Table 5.2: Number of possible TSL$_2$ grammars for $\Sigma = \{$s, ʃ, p, t, a$\}$

$\lvert T\rvert$	# possible T $= \binom{\lvert\Sigma\rvert}{\lvert T\rvert}$	# possible R $= 2^{\lvert T\rvert^2}$	# possible grammars = # of $T \times$ # of R
5	1	33 554 432	33 554 432
4	5	66 536	332 680
3	10	512	5 120
2	10	16	160
1	5	2	10
0	1	1	1
			Total = 33 892 403

As a general note on the learnability of TSL$_k$ languages, I point out that for any finite value of k and a Σ with finitely many segments, the number of possible grammars may be extremely large, but is nonetheless finite. As such, for the class of TSL$_2$ languages, even an algorithm that picks among grammars randomly will be successful in the limit, thus achieving Gold-learnability (Gold, 1967). However, when Heinz et al. (2011) first proposed the TSL class of formal languages, they investigated a number of computational properties of the class, but did not know whether or not it was possible to learn a TSL$_k$ grammar *efficiently* (i.e. with polynomial bounds on time and data; de la Higuera, 1997) without prior knowledge of the segments in T. Following up on the issue, Jardine and Heinz (2015) provide an algorithm that does so (see also Jardine, 2015). The algorithm, when given sufficient data in the training set, simultaneously acquires the two components of the grammar (T and R) in an efficient manner for any TSL$_2$ language, regardless of the size

[1]To further press the issue I point out that the number of possible TSL$_2$ grammars for a relatively modest inventory of 20 segments is on the order of 10^{120}. For comparison, the number of atoms in the universe is estimated to be on the order of 10^{80}.

of Σ. An outline of the technical aspects of this algorithm (named the Tier-based Strictly 2-Local Inference Algorithm, or 2TSLIA, by Jardine and Heinz, 2015) is provided below in Section 5.2.1, though I point out that my summary sacrifices a certain amount of mathematical precision in favour of accessibility to a broader readership. Jardine and Heinz (2015) do not ignore these details, and the reader is referred to their article for a full description of the 2TSLIA as well as formal proofs of certain mathematical properties of the learner. Section 5.2.2 shows how the 2TSLIA would learn a pattern of sibilant harmony that is blocked by coronal obstruents (similar to the case of Slovenian, which was presented in example (4) in Chapter 4)—an example that illustrates the practical importance of several aspects of the learning algorithm.

5.2.1 Summary of the Tier-based Strictly Local Inference Algorithm

Of crucial importance to the 2TSLIA is the concept of a 2-*path* (Jardine and Heinz, 2015). A 2-path is formally defined as a 3-tuple $\langle x, Z, y\rangle$ where x and y are segments in Σ that occur in a particular word (potentially two instances of the same segment), and Z is the set of segments that intervene between x and y in the word (note that Z includes only one instance of the same intervening segment, and thus Z is a subset of Σ). For example, in the word [sapʃ], [s] precedes [ʃ] and the two segments are separated by intervening segments [a] and [p]. The corresponding 2-path is ⟨s, {a, p}, ʃ⟩. Likewise, the entire set of 2-paths for the word [sapʃ] includes each of the following: ⟨s, {}, a⟩, ⟨s, {a}, p⟩, ⟨s, {a, p}, ʃ⟩, ⟨a, {}, p⟩, ⟨a, {p}, ʃ⟩, ⟨p, {}, ʃ⟩. Intuitively, 2-paths can therefore be thought of as $x \dots y$ precedence relations that are augmented with the set of segment interveners.

The 2TSLIA takes as its input a segment inventory with each segment labelled for some (arbitrary) order, $\Sigma = \{\sigma_1, \sigma_2, \sigma_3, \dots, \sigma_n\}$, as well as a finite set I of grammatical strings (i.e. words that adhere to the phonotactic restrictions of the target language) that serves as the algorithm's training data. The output of the learning algorithm is a TSL_2 grammar defined by $G = \langle T, S\rangle$ (with a grammar in the form $G = \langle T, R\rangle$ easily obtained as well).

The algorithm begins with a hypothesized tier $T_0 = \Sigma$ (i.e. every segment in the inventory is on the tier) and attempts to remove segments from the tier one at a time (iteratively hypothesizing a potentially different tier in each step: T_0, T_1, …, T_n). In order to safely remove a segment σ_i from T_{i-1} (such that in the next step $T_i = T_{i-1} - \{\sigma_i\}$), two conditions need to be satisfied. First, σ_i must be what Jardine and Heinz (2015) call a *free element*. The second condition is that σ_i is not what they call an *exclusive blocker*. Being a free element means that σ_i may freely co-occur with any segment (including itself), both preceding and following it. This amounts to stating that σ_i would not be present in any of the 2-factors in R if T_{i-1} (the current guess of T) is in fact the correct tier. To determine whether or not this is true in the first step, where $T_0 = \Sigma$, the algorithm checks the set of 2-paths that are present in the training data. If for every segment $\sigma' \in \Sigma$, the set of 2-paths includes both $\langle \sigma_i, \{\}, \sigma' \rangle$ and $\langle \sigma', \{\}, \sigma_i \rangle$, then σ_i satisfies the first (free element) condition. Note that in subsequent iterations of the algorithm (when certain segments have been removed from the hypothesis for T), 2-paths of the form $\langle \sigma_i, \{Z\}, \sigma' \rangle$ and $\langle \sigma', \{Z\}, \sigma_i \rangle$, where Z contains only segments that have already been removed from the tier (i.e. $Z \subseteq \Sigma - T_{i-1}$), may also provide evidence that σ_i is a free element.

If the free element condition is satisfied for σ_i, the algorithm then moves on to the second (exclusive blocker) condition. Being an exclusive blocker means that the segment is a member of the tier T, but is not present in any member of the 2-factor restrictions in R. In practice, this means that it can block a phonotactic dependency, even though it does not actively participate in the restrictions, such as a [t] that blocks sibilant harmony. More formally, having determined that σ_i is a free element, we also know σ_i is not part of any 2-factor in R. However, this is not enough evidence to safely remove σ_i from the tier. That is, if σ_i is indeed a member of the tier T that is specified by the target grammar G, then there may be some 2-factor $*\sigma_x\sigma_y \in R$ such that if σ_i was incorrectly removed from T, then $*\sigma_x\sigma_y$ would erroneously be evidenced in a tier-adjacent context in the input data. Specifically, this type of case would arise if there is a grammatical word $\sigma_x\sigma_i\sigma_y$—removing σ_i from the tier may cause the learner to determine that σ_x or σ_y is a free element, even though they are actually part of a 2-factor restriction in the target grammar (and hence both must be in T). In order to determine if σ_i is an exclusive blocker, the algorithm searches the set of 2-paths to ensure that there is no pair of segments, σ_x and σ_y, whose tier-based adjacency is dependent on the presence of an intervening σ_i (see Step 4 in Section 5.2.2 for a practical illustration).

If σ_i satisfies both of the above conditions (i.e. it is a free element, but not an exclusive blocker), then then learner's hypothesized tier T_{i-1} is updated to T_i with the removal of σ_i. If either of the conditions is not satisfied, then no change is made and $T_i = T_{i-1}$. The entire process is then repeated for $\sigma_{i+1}, \sigma_{i+2}, \ldots, \sigma_n$, at which point the learner will have determined whether or not each of the segments in the inventory is a member of *T*.

Once the 2TSLIA has induced the tier *T*, the final step is to fill in *S* with the set of tier-based 2-factors that are observed in the data. To do this, the learner can simply check each of the 2-paths $\langle x, Z, y \rangle$. If none of the segments in *Z* are also members of *T*, then *x* and *y* are tier-adjacent, and the 2-factor *xy* is added to *S*. If the desired output is $G = \langle T, R \rangle$, then *R* can be obtained by taking the complement of *S* with respect to all possible 2-factors comprised of segments in *T*.

5.2.2 Example of the 2TSLIA: Sibilant Harmony with Blocking

Target pattern In the following example, I show each step of the 2TSLIA to illustrate how it would learn a case of sibilant harmony that bans *[s…ʃ] and *[ʃ…s] subsequences, unless a coronal obstruent [t] intervenes. With an inventory of $\Sigma =$ {a, s, ʃ, t, p}, the target grammar is thus $G = \langle T = \{$s, ʃ, t$\}, S = \{$ss, st, ʃʃ, ʃt, ts, tt, tʃ$\}\rangle$, or equivalently $G = \langle T = \{$s, ʃ, t$\}, R = \{$*sʃ, *ʃs$\}\rangle$.

Input to 2TSLIA The algorithm is provided with a segment inventory $\Sigma = \{$a, s, ʃ, t, p$\}$, which has each segment labelled in some (potentially arbitrary) order. In this case, let $\sigma_1 =$ [a], $\sigma_2 =$ [s], $\sigma_3 =$ [ʃ], $\sigma_4 =$ [t], and $\sigma_5 =$ [p]. The algorithm is also provided with a set of grammatical training words *I*. The types of words that need to be included (or not included) in *I* will be discussed throughout the example and aggregated in (16) at the end of this section as a sufficient set of training items.

Step 1: determining that [a] is not in *T* The learner's initial hypothesis is that the tier is $T_0 = \Sigma = \{$a, s, ʃ, t, p$\}$. Beginning with the segment [a] (since it is labelled σ_1), the learner must determine whether it meets the free element condition, and if so, whether it meets the exclusive blocker condition. As we already know that

[a] is not a member of the tier specified by the target grammar, we know that it must be a free element but not an exclusive blocker. In order for the 2TSLIA to determine this as well, it proceeds as follows. First, it looks for each of the nine possible 2-paths of the form $\langle \text{a}, Z, \tau \rangle$ and $\langle \tau, Z, \text{ a} \rangle$ where τ is any member of the current guess $T_0 = \{\text{a, s, ʃ, t, p}\}$, and Z does not include any members of T_0. Since $T_0 = \Sigma$, it is therefore looking for evidence that [a] may occur in a string-adjacent position with all segments. The following two members of the input strings I, which are both grammatical words in the target language, would be sufficient for it to do so: {tasapa, ʃaataʃ}. The relevant subset of 2-paths provided by [tasapa] includes ⟨t, {}, a⟩, ⟨a, {}, s⟩, ⟨s, {}, a⟩, ⟨a, {}, p⟩, ⟨p, {}, a⟩, and the remaining four possibilities, ⟨ʃ, {}, a⟩, ⟨a, {}, a⟩, ⟨a, {}, t⟩, ⟨a, {}, ʃ⟩, are provided as a subset of the 2-paths in [ʃaataʃ]. The segment [a] is therefore a free element, and the algorithm moves on to the next step.

In order for the algorithm to determine that [a] is not an exclusive blocker, and that it may be safely removed from the tier, the input training items must include evidence that no pair of segments requires [a] to intervene in order for them to co-occur. To state this in a more intuitive way, the set of segment pairs x, y that are observed in 2-paths of the form $\langle x, \{\}, y \rangle$ must not be a proper subset of the segment pairs that are observed in $\langle x, \{\text{a}\}, y \rangle$ 2-paths (where neither x nor y is [a]). Note that this condition can easily be satisfied since all string-adjacent 2-factors except {*sʃ, *ʃs} are permitted, and I might include, for example, {assa, asta, aspa, aʃʃa, aʃta, aʃpa, atsa, atʃa, atta, atpa, apsa, apʃa, apta, appa}. Finally, since words containing *[…saʃ…] or *[…ʃas…] are prohibited, the input strings in I will not include any instances of ⟨s, {a}, ʃ⟩ or ⟨ʃ, {a}, s⟩. With this type of evidence the 2TSLIA can determine that [a] is not an exclusive blocker and that it can be safely removed from the tier for the algorithm's next hypothesis.

Steps 2 and 3: [s] and [ʃ] are in *T* After removing [a] from the tier, the algorithm's next step is to hypothesize a tier $T_1 = \{\text{s, ʃ, t, p}\}$, and to check whether $\sigma_2 = $ [s] should also be removed. It is easy to see, however, that [s] is not a free element in the language. Since the language prohibits any *[…s(a)ʃ…] or *[…ʃ(a)s…] sequences, the training set I will contain no instances of the 2-paths ⟨s, {}, ʃ⟩, ⟨s, {a}, ʃ⟩ or ⟨ʃ, {}, s⟩, ⟨ʃ, {a}, s⟩, which would be required in order for

[s] to satisfy the free element condition. Since [s] is not a free element, the learner does not need to check the exclusive blocker condition, and instead moves directly to the next hypothesis: $T_2 = \{$s, ʃ, t, p$\}$. The same result will hold during the 2TSLIA's subsequent iteration for $\sigma_3 =$ [ʃ], and thus both [s] and [ʃ] remain in the learner's guess of T. (Note that no changes to the hypothesized tier were made in these steps, and $T_1 = T_2 = T_3 = \{$s, ʃ, t, p$\}$.)

Step 4: [t] is an exclusive blocker With $T_3 = \{$s, ʃ, t, p$\}$, the learner moves on to check the two conditions that determine whether or not $\sigma_4 =$ [t] is in T. Since [t] is not present in any member of R, we know that it is a free element and so it will satisfy the first condition. The evidence that is necessary for the algorithm to conclude this is available from the members of I already provided in a previous step: {asta, aʃta, atsa, atʃa, atta, atpa, apta}. This set of training items results in each of the nine 2-paths that are required for [t] to be a free element: ⟨a, {}, t⟩, ⟨t, {}, a⟩, ⟨s, {}, t⟩, ⟨ʃ, {}, t⟩, ⟨t, {}, s⟩, ⟨t, {}, ʃ⟩, ⟨t, {}, p⟩, ⟨p, {}, t⟩. (In fact, the algorithm does not technically need to search for the first two members of this list of 2-paths, as [a] was already removed from the tier after Step 1.)

Since [t] satisfies the free element condition, the 2TSLIA also checks the second condition, which in this case will determine that [t] is an exclusive blocker. The reason for this is as follows. We know that the grammatical words in I will not contain any instances of the 2-paths ⟨s, {}, ʃ⟩ or ⟨ʃ, {}, s⟩, ⟨s, {a}, ʃ⟩ or ⟨ʃ, {a}, s⟩. However, since the pattern of sibilant harmony is blocked by intervening coronal obstruents, if there is even one instance of a word such as [astʃa] or [ʃatasa], which include the 2-paths ⟨s, {t}, ʃ⟩ and ⟨ʃ, {a, t}, s⟩, the algorithm will conclude that [t] is an exclusive blocker and must be a member of the tier T. Specifically, this is because the presence of a ⟨ʃ, {Z}, s⟩ 2-path, where Z is a subset of {a, t} is dependent on the presence of [t] in Z.

Step 5: [p] is not in *T* The hypothesized tier is now $T_4 = \{$s, ʃ, t, p$\}$ (this has not changed since the end of Step 1), and the last of the five segments in the inventory to be checked is $\sigma_5 =$ [p]. It is clear that [p] is a free element, as the example words in I provided above include [p] in a string-adjacent context with all members of the inventory, including itself. Unlike [t], however, [p] is not an exclusive blocker since any words of the form *[aspʃa] or *[ʃapasa] would violate the phonotactics of the TSL_2 language, and therefore could not be provided in I. Since [p] is a free element but not an exclusive blocker, the algorithm may safely remove it from the tier, arriving at its final (correct) hypothesis that $T_5 = T = \{\text{s}, \text{ʃ}, \text{t}\}$.

Step 6: 2-factors in *S* (or *R*) After running through each $\sigma_i \in \Sigma$, and having discovered the correct tier T that is specified by the grammar, the last step that the 2TSLIA needs to complete is to compile a list of all tier-based 2-factors that are observed in I. To achieve this, the algorithm simply records all pairs τ_1, τ_2 where $\tau_1, \tau_2 \in T$ that are observed in 2-paths of the form $\langle \tau_1, Z, \tau_2 \rangle$, where Z does not include any members of T. This will result in the correct specification of $S = \{$ss, st, ʃʃ, ʃt, ts, tt, tʃ$\}$, which completes the learning of the target TSL_2 grammar for this example. If the preferred format of the grammar is in terms of the tier-based 2-factor restrictions (i.e. the 2-factors in R), we can simply take the complement of S with respect to the set of all possible 2-factors comprised of segments in T. In the present case, this yields $R = \{$*sʃ, *ʃs$\}$.

Representative list of training words The set of words was used in the above example of the 2TSLIA (Jardine and Heinz, 2015) was as follows:

(16) $$I = \left\{ \begin{array}{c} \text{tasapa, ʃaataʃ, assa, asta, aspa, aʃʃa, aʃta, aʃpa, atsa,} \\ \text{atʃa, atta, atpa, apsa, apʃa, apta, appa, astʃa, ʃatasa} \end{array} \right\}$$

Crucially, the training words must include a set of 2-paths that provides sufficient evidence about whether each segment is a free element and whether it is an exclusive blocker. That is, if a particular 2-path is absent from the training data, the algorithm runs the risk of leaving a segment on the hypothesized tier either because it induces that it is a not free element (when there are in fact no relevant phonotactic restrictions on its occurrence) or that it is an exclusive blocker (even if it does not actually block any dependencies).

5.2.3 Conclusion: The TSL_2 Class Is Learnable

When Heinz et al. (2011) proposed the class of Tier-based Strictly 2-Local formal languages, they did not know whether or not it would be possible to design an algorithm that could efficiently learn the entire class without prior knowledge of the contents of T. The 2TSLIA designed by Jardine and Heinz (2015), which is provably efficient and correct, thus provides a significant increase in the plausibility of a model that equates the human learner's hypothesis space to the class of TSL_2 stringsets, as the algorithm offers a computationally tractable solution to the problem of TSL_2 learnability.

5.3 Is the TSL_2 Region Too Small?

As a final question about the empirical validity of the TSL_2 approach, I consider whether or not there are any patterns beyond the proposed TSL_2 boundary that should also be included in the set of possible, human-learnable languages. In what follows, I describe a number of languages whose overall (consonant) phonotactics cannot be generated with a single TSL_2 grammar. However, each of the cases presented is clearly composed of multiple, and in some cases interacting, TSL_2 dependencies that can be characterized individually before combining them. I argue that the theory needs to be expanded in order to accommodate such systems in a mathematically principled fashion, and I discuss a few preliminary suggestions for how we might do so.

5.3.1 Multiple Non-Conflicting TSL_2 Patterns

The first type of phonotactic system to consider is one whose grammar must specify two different sets of restrictions (R_1 and R_2) indexed for two different tiers (T_1 and T_2). To illustrate the problem, I present data from two different Berber languages: Tamashek Tuareg (Heath, 2005; Hansson, 2010a; Bennett, 2013) and Imdlawn Tashlhiyt (Elmedlaoui, 1995; Hansson, 2010a,b). Each of the patterns highlights a number of interesting issues that must be taken into account as we extend the theory to account for more complex phonotactic patterns.

5.3.1.1 Tamashek Tuareg: Two Long-Distance Dependencies

The Tamashek dialect of Tuareg (Berber; Heath, 2005) has a pattern of sibilant harmony similar to a what is found in many other Berber languages. The basic generalization, illustrated in (17) with alternations of the causative prefix /s(:)-/, is that sibilants must agree in both anteriority and voicing. (17) shows that the prefix surfaces faithfully as [s-] when the root contains no sibilants, or when the only sibilant in the root is [s]. (18) demonstrates anticipatory harmony, in that the prefix is required to agree in both anteriority and voicing with any other sibilant [ʃ, z, ʒ] that is present in the root. Note that the surface form of the vowels in Tamashek Tuareg varies considerably by context, and the below examples of causative verbs simply use 'V' to represent any vowel if a full surface form is not provided in Heath (2005) (note that Heath further distinguishes between 'short' and 'full' vowels).

(17) Tamashek Tuareg: underlying causative prefix /s-/ (Heath, 2005)

	Causative	*Gloss*
a.	-**s**-VdufV-	'make plump'
b.	-**s**-VŋŋV-	'cook'
c.	-**s**-V**s**VfVr-	'treat (patient)'
d.	-**s**-V**s**kVr -	'hold upright'

(18) Tamashek Tuareg: unbounded sibilant harmony (Heath, 2005)

	Causative	*Gloss*
a.	-**ʃ**-VlVjtV**ʃ**-	'shake off'
b.	**ʃ**-ùkmə**ʃ**	'make scratch!'
c.	-**z**-Vg**z**Vl-	'shorten'
d.	**z**ˤ-ìhə**z**ˤ	'make approach!'
e.	-**ʒ**-V**ʒ**VlwVʁ-	'glare at'

The above pattern of sibilant harmony is easily represented with the following TSL_2 grammar: $G_1 = \langle T_1 = \{\text{s}, \text{ʃ}, \text{z}, \text{ʒ}\}, R_1 = \{\text{*sʃ, *sz, *sʒ}\}\rangle$.[2]

[2] G_1 is slightly simplified. Heath (2005) points out that the restriction on co-occurring sibilants seems to hold more generally, even morpheme-internally and so a full set of 2-factors restrictions would be $R_1 = \{$*sʃ, *sz, *sʒ, *ʃs, *ʃz, *ʃʒ, *zs, *zʃ, *zʒ, *ʒs, *ʒʃ, *ʒz$\}$.

Interestingly, however, the sibilant harmony co-exists with an independent pattern of long-distance labial dissimilation. This is exemplified below with /m/→[n] alternations in several types of prefix, including the medio-passive and the agentive, when they are followed by a labial [b, f, m] at any distance.

(19) Tamashek Tuareg: underlying prefixes containing /m/ (Heath, 2005)

a.	æ-m-ɑ́jrɑd	'one who can disappear'	(agentive)
b.	-æ̀m-erɑ-	'be opened (Perf)'	(medio-passive)

(20) Tamashek Tuareg: long-distance labial dissimilation (Heath, 2005)

a.	-ə̀nː-ə**b**dˤɑ-	'be dislocated'	*-ə̀**m**ː-ə**b**dˤɑ-
b.	ɑ-n-ə̀**f**rən	'be chosen'	*ɑ-**m**-ə̀**f**rən
c.	ɑ-n-ɑ́nɑ**m**	'one who is fond'	*ɑ-**m**-ɑ́nɑ**m**

The pattern of long-distance labial dissimilation can likewise be generated with a TSL$_2$ grammar: $G_2 = \langle T_2 = \{\text{b, f, m}\}, R_2 = \{\text{*mb, *mf, *mm}\}\rangle$.

As it stands, Tamashek Tuareg exhibits two independent long-distance dependencies that can individually be characterized as TSL$_2$ patterns. However, since the two grammars specify two completely different tiers, we must ask whether or not they can be combined into a single grammar that requires just one tier with a single set of restrictions. One way of attempting to achieve this might be to take the union of the tiers ($T_1 \cup T_2$), and the union of the sets of 2-factor restrictions ($R_1 \cup R_2$), resulting in $G_3 = \langle T_3 = \{\text{s, ʃ, z, ʒ, b, f, m}\}, R_3 = \{\text{*sʃ, *sz, *sʒ, *mb, *mf, *mm}\}\rangle$. This strategy results in the incorrect prediction that the set of labial consonants should block sibilant harmony, and that the set of sibilants should block labial dissimilation. For example, since both of the 2-factors [sm] and [mʃ] are permitted on T_3, we would erroneously allow a word such as *[**s**-ùk**m**ɑ**ʃ**] (cf. [ʃ-ùkməʃ] in (18b)).

As an alternative, I point out that in cases where patterns can individually be described in TSL$_2$ terms with no overlap in the members of each tier (and hence no overlap in the tier-based 2-factor restrictions), a single grammar can be generated as the *conjunction* of each of the TSL$_2$ grammars. That is, a string is a member of the language if and only if it is permitted by each of the conjoined grammars, adhering to all of the individual patterns. This is done below with the grammar in

(21).

(21) $$G = \left\langle \begin{array}{c} T_1 = \{\text{s, ʃ, z, ʒ,}\}, \\ R_1 = \{\text{*sʃ, *sz, *sʒ}\} \end{array} \right\rangle \wedge \left\langle \begin{array}{c} T_2 = \{\text{b, f, m}\}, \\ R_2 = \{\text{*mb, *mf, *mm}\} \end{array} \right\rangle$$

As a useful test case for the above grammar, Heath (2005) does provide one example of the two patterns being upheld in a single word, which is given in (22).

(22) Simultaneous sibilant harmony and labial dissimilation (Heath, 2005)
ɑ-**z**ˤ-ə**n**ː-ət-ə́l**m**ə**z**ˤ ‘act of spitting up saliva’ *ɑ-**s**ˤ-ə**m**ː-ət-ə́l**m**ə**z**ˤ

5.3.1.2 Imdlawn Tashlhiyt: Sibilant Harmony With Partial Blocking

The Imdlawn dialect of Tashlhiyt (Berber; Elmedlaoui, 1995; Hansson, 2010a,b) has a pattern of sibilant harmony nearly identical to what was illustrated above in (17) and (18) for Tamashek Tuareg. The basic generalization, again demonstrated with alternations of the causative prefix /s(ː)-/, is that sibilants must agree in both anteriority and voicing. (23) shows that the prefix surfaces as [s-] when no sibilants follow it, or when the only sibilant in the root is [s]. In (24), the prefix agrees in both anteriority and voicing with any other sibilant [ʃ, z, ʒ] that is present in the root.

(23) Imdlawn Tashlhiyt: underlying causative prefix /s-/
(Elmedlaoui, 1995; Hansson, 2010a,b)

	Base	*Causative*	*Gloss*
a.	gdʷm	**s**-gdʷm	‘arrange upside down’
b.	uga	**s**ː-uga	‘be evacuated’
c.	n**s**a	**s**ː-n**s**a	‘spend the night’
d.	a**s**ːtwa	**s**-a**s**ːtwa	‘settle, be levelled’

(24) Imdlawn Tashlhiyt: sibilant harmony
(Elmedlaoui, 1995; Hansson, 2010a,b)

	Base	*Causative*	*Gloss*
a.	fia**ʃ**r	**ʃ**-fia**ʃ**r	‘be full of straw, of discord’
b.	bːuk**ʃ**ːa	**ʃ**-buk**ʃ**ːa	‘be full to overflowing’
c.	bru**z**ːa	**z**-bru**z**ːa	‘crumble’
d.	n**z**a	**z**ː-n**z**a	‘be sold’
e.	mː**ʒ**dawl	**ʒ**-mː**ʒ**dawl	‘stumble’
f.	gˤrˤuˤ**ʒ**ˤːmˤ	**ʒ**ˤ-gˤrˤuˤ**ʒ**ˤːmˤ	‘be extinguished (in cooking)’

The phonotactics of the basic form of Imdlawn Tashlhiyt sibilant harmony shown in (23) and (24) are easily captured as a TSL_2 language, using a formal grammar that is typical of a pattern with unbounded locality—the tier *T* contains exactly the set of segments that actively participate in the dependency, namely the sibilants in this case, and the set of permitted 2-factors on the tier is $S = \{$ss, ʃʃ, zz, ʒʒ$\}$ (with *R* containing all other permutations of two sibilants). This would be enough to characterize the pattern of sibilant harmony as it is found in other varieties of Berber, but the case of Imdlawn Tashlhiyt is peculiar in that the requirement for agreement in voicing (but not anteriority) is blocked by intervening voiceless obstruents. This is shown below in (25), with examples for each of the voiceless obstruents [ħ, k, f, χ, q] blocking voicing harmony, and with (25e)-(25f) demonstrating that agreement for anteriority is still enforced across a blocker.

(25) Imdlawn Tashlhiyt: sibilant voicing harmony blocked
(Elmedlaoui, 1995; Hansson, 2010a,b)

	Base	*Causative*	*Gloss*
a.	**ħuz**	**s-ħuz**	‘annex’
b.	u**kz**	**s**ː-u**kz**	‘recognize’
c.	rˤuˤ**f**ˤː**z**ˤ	**s**ˤ-rˤuˤ**f**ˤ**z**ˤ	‘appear resistant, recalcitrant’
d.	m-**χ**a**z**aj	**s**m**χ**a**z**aj	‘loathe each other’
e.	**q**ːu**ʒ**ːi	**ʃ-q**u**ʒ**ːi	‘be dislocated, broken’
f.	mˤ-**ħ**ˤaˤrˤaˤ**ʒ**ˤ	**ʃ**ˤ-mˤ-**ħ**ˤaˤrˤaˤ**ʒ**ˤ	‘get angry with each other’

In contrast to the similar, previously described cases of sibilant harmony with blocking that occur in Slovenian (see Section 4.2.1) and in Kinyarwanda (see Section 5.1.1), the pattern in Imdlawn Tashlhiyt cannot be captured using a TSL$_2$ grammar whose tier includes all of the sibilants and all of the blockers. With respect to the above data, this would yield $T_1 = \{$s, ʃ, z, ʒ, ħ, k, f, χ, q$\}$, and the observed set of 2-factors on this tier is $S_1 = \{$ss, ʃʃ, zz, ʒʒ, sħ, ħz, sk, kz, sf, fz, sχ, χz, ʃq, qʒ, ʃħ, ħʒ$\}$. However, this cannot be the correct grammar for Imdlawn Tashlhiyt—the presence of both $\{$sħ, ħʒ$\}$ on the relevant tier falsely implies the grammaticality of a word such as *[**s**ˤ-mˤ-**ħ**ˤaˤrˤaˤ**ʒ**ˤ] (cf. the correct form in (25f) [**ʃ**ˤ-mˤ-**ħ**ˤaˤrˤaˤ**ʒ**ˤ]). However, if we leave out the voiceless obstruents, with $T_2 = \{$s, ʃ, z, ʒ$\}$, the data above exhibit each of the tier-based 2-factors in $S_2 = \{$ss, ʃʃ, zz, ʒʒ, sz, ʃʒ$\}$, misidentifying *[**s**-bru**z**:a] and *[**ʃ**-m:**ʒ**dawl] as permissible strings (cf. (24c) and (24e) above).

Instead we can characterize the above phonotactics as a combination of two patterns of agreement (one for anteriority, one for voicing) using a conjunction of TSL$_2$ languages, as was done for Tamashek Tuareg in the previous section. The difference for the present case is that here the two patterns operate on tiers that partially overlap. One applies on a tier that is made up of all sibilants and the voiceless obstruents $T_1 = \{$s, ʃ, z, ʒ, ħ, k, f, χ, q$\}$, banning 2-factors whose two members are sibilants that disagree in voicing $R_1 = \{$*sz, *sʒ, *ʃz, *ʃʒ, *zs, *zʃ, *ʒs, *ʒʃ$\}$. The second pattern is enforced on the tier of sibilants $T_2 = \{$s, ʃ, z, ʒ$\}$, banning any 2-factors of sibilants that disagree for anteriority $R_2 = \{$*sʃ, *sʒ, *ʃz, *ʃs, *zʃ, *zʒ, *ʒs, *ʒz$\}$. A grammar that conjoins these two patterns would be sufficient to characterize the phonotactics of Imdlawn Tashlhiyt as presented above, and is given below in (26).

(26)
$$G = \begin{array}{c} \left\langle T_1 = \left\{ \begin{array}{c} \text{s, ʃ, z, ʒ,} \\ \text{ħ, k, f, χ, q} \end{array} \right\}, R_1 = \left\{ \begin{array}{c} \text{*sz, *sʒ, *ʃz, *ʃʒ,} \\ \text{*zs, *zʃ, *ʒs, *ʒʃ} \end{array} \right\} \right\rangle \\ \bigwedge \\ \left\langle T_2 = \{\text{s, ʃ, z, ʒ}\}, R_2 = \left\{ \begin{array}{c} \text{*sʃ, *sʒ, *ʃz, *ʃs,} \\ \text{*zʃ, *zʒ, *ʒs, *ʒz} \end{array} \right\} \right\rangle \end{array}$$

Lastly, there are a couple of interesting notes about the grammar in (26). First is that T_1 is another example of a tier that cannot be classified as a natural class, since it includes *all sibilants* (both voiced and voiceless), but only the *voiceless obstruents* (voiced obstruents are not blockers), and so this pattern serves as further support for not restricting the set of tiers that a learner may consider to natural classes (see Section 5.1 for more examples). Second, note not only that T_2 is a subset of T_1 but also that several 2-factors occur both in R_1 and R_2, namely each of {*sʒ, *ʃz, *zʃ, *ʒs}. This redundancy does not affect the success of the grammar, but it does leave room for a more efficient grammar. Specifically, the grammar needs only to specify these 2-factors restrictions for the sibilant tier (i.e. T_2 in this case), since the absence of a 2-factor on a tier implies its absence on any superset of that tier. We could therefore reduce R_1 to {*sz, *ʃʒ, *zs, *ʒʃ}, since the banning of {*sʒ, *ʃz, *zʃ, *ʒs} is enforced on T_2 (implying that they may not occur on T_1 either). Further issues concerning the overlap of the formal properties of co-existing patterns are certainly of broader interest, but are left for future research.

5.3.2 Multiple Conflicting TSL$_2$ Dependencies

With the types of patterns like those of Tamashek Tuareg and Imdlawn Tashlhiyt as motivation for further study of multiple TSL$_2$ patterns, McMullin and Allen (2015) offered a preliminary investigation of the computational properties of TSL$_2$ conjunctions, arguing that they form a lattice class of languages (Heinz et al., 2012), and are therefore Gold-learnable (Gold, 1967). However, there also exist several attested patterns that cannot be described as a conjunction of TSL$_2$ patterns, since there are multiple individual dependencies between consonants (each characterizable as TSL$_2$) that are in direct conflict with each other.

5.3.2.1 Revisiting Tamashek Tuareg

As a first example, I expand on the phonotactic generalizations of Tamashek Tuareg that were presented above in Section 5.3.1.1. Recall that the language includes two separate long-distance dependencies with unbounded locality: sibilant harmony and labial dissimilation. The present discussion is relevant only for the latter of these patterns, for which the data are reproduced below in (27) and (28), in which a prefix containing /m/ surfaces with [n] when a labial consonant [b, f, m] follows it at any distance. All data are cited from (Heath, 2005).

(27) Tamashek Tuareg: underlying prefixes containing /m/

	a.	æ-m-ɑ́jrɑd	'one who can disappear'	(agentive)
	b.	-æ̀m-erɑ-	'be opened (perf.)'	(medio-passive)

(28) Tamashek Tuareg: long-distance labial dissimilation

	a.	-ə̀nː-ə**b**dˤɑ-	'be dislocated'	*-ə̀**m**ː-ə**b**dˤɑ-
	b.	ɑ-n-ə̀**f**rən	'be chosen'	*ɑ-**m**-ə̀**f**rən
	c.	ɑ-n-ɑ́nɑ**m**	'one who is fond'	*ɑ-**m**-ɑ́nɑ**m**

A TSL_2 grammar for the pattern of labial dissimilation is provided in (29), which accounts for the generalization that prefixes containing /m/ do not surface faithfully when they precede a labial consonant at any distance.

(29) $G = \langle T = \{\text{b, f, m}\}, R = \{\text{*mb, *mf, *mm}\}\rangle$

However, there is also evidence that the same prefixes are permitted to surface faithfully if the resulting [m] precedes an *adjacent* [b], as shown in (30).

(30) Tamashek Tuareg: adjacent nasal place assimilation resulting in [mb]

	a.	-æ̀m-bæbbɑ-	'carried each other (perf.)'	*-æ̀n-bæbbɑ-
	b.	-æm-bə̀lədˤwəj-	'fell over (perf.)'	*-æn-bə̀lədˤwəj-

The fact that [mb] sequences are permitted directly contradicts the grammar in (29). A general property of TSL_2 languages is that if $T_x \subseteq T_y$ and a 2-factor $*\sigma_1\sigma_2$ is banned on T_x, then $*\sigma_1\sigma_2$ is also banned on T_y. Intuitively, this translates to the idea that an unbounded dependency should also hold in transvocalic contexts, string-adjacent contexts, and so on. However, the data above show that there is a restriction against *[mb] on $T = \{\text{b, f, m}\}$, even though [mb] is permitted when $T = \Sigma$ (i.e. when they are adjacent on a tier that includes all segments). Since $\{\text{b, f, m}\}$ is a subset of the segment inventory Σ, it is not possible to achieve independent TSL_2 characterizations of each pattern without contradicting the other.

The resulting problem—that one pattern is violated in favour of another—is not new to phonologists. From a formal language theoretic perspective, an ideal solution might be to define a new mathematical operator (i.e. in addition to conjunction, disjunction, etc.) that would allow for this type of combination of individual formal grammars. From the perspective of theoretical phonology, there are already (at least) two well-known strategies for solving the problem: rule ordering and constraint ranking (or weighting). I suggest that the best solution is to circle back to a constraint-based approach (e.g. OT; Prince and Smolensky, 2004), representing each constraint against the co-occurrence (adjacent or non-adjacent) of two segments as individual, violable and ranked TSL_2 grammars. An example of this is shown below in Tableaux (31) and (32), each of which compares two possible surface strings with [m/n] preceding [b] at some distance.

(31) Tamashek Tuareg: non-adjacent disagreement

	$\left\langle \begin{array}{c} T = \Sigma \\ R = \{\text{*nb}\} \end{array} \right\rangle$	$\left\langle \begin{array}{c} T = \{\text{b, f, m}\} \\ R = \{\text{*mb, *mf, *mm}\} \end{array} \right\rangle$
a. -ə̀mː-əbdˤɑ-		$*!$
b. ☞ -ə̀nː-əbdˤɑ-		

(32) Tamashek Tuareg: adjacent agreement

	$\left\langle \begin{array}{c} T = \Sigma \\ R = \{\text{*nb}\} \end{array} \right\rangle$	$\left\langle \begin{array}{c} T = \{\text{b, f, m}\} \\ R = \{\text{*mb, *mf, *mm}\} \end{array} \right\rangle$
a. ☞ -æ̀m-bæbbɑ-		$*$
b. -æ̀n-bæbbɑ-	$*!$	

Note that (31) and (32) are meant to provide a visualization of how formal grammars can be thought of as independent markedness constraints in an Optimality Theory framework. As I continue to limit the scope of my dissertation to phonotactics in particular, the tableaux do not include any input forms or faithfulness constraints—the intention is to illustrate the relative preference between two possible output forms. Other, potentially optimal output candidates that do not violate any markedness constraints are presumed to be ruled out for independent reasons. Another example of a pattern that requires a similar approach is provided in the next section, prior to continuing the discussion of TSL_2 constraints in Section 5.3.3.

5.3.2.2 Samala Sibilant Harmony Overrides Palatalization

Though it may seem intuitive that a restriction against two adjacent segments trumps a dependency that holds at longer distances, this is not always the case. As empirical evidence in support of this, I present data from Samala (Ineseño Chumash; Applegate, 1972), in which a regressive sibilant harmony with unbounded locality overrides a restriction against string-adjacent {*st, *sn, *sl} that results in a pattern of dissimilation (in which, e.g., /st/ surfaces as [ʃt]).

This is a well-known and often-cited case in the theoretical literature on long-distance interactions in phonology, starting with Poser (1982). It is important to note that in that body of theoretical works, the descriptive generalization has typically been understood as being the exact opposite, with local dissimilation (palatalization before /t,n,l/) overriding the non-local sibilant harmony (Poser, 1982, 1993; McCarthy, 2007; Hansson, 2010a). However, closer scrutiny of the primary descriptive source, Applegate's (1972) grammar, strongly suggests that he intends to describe the interaction of these patterns as it is presented below.[3] All data in this section are drawn directly from Applegate (1972).

The data in (33) illustrate the pattern of anticipatory sibilant harmony, in which sibilants are required to agree in anteriority with any sibilant that follows it. Note that both [–ant] and [+ant] segments can be triggers or targets, and that the dependency holds across relatively large distances.

(33) Samala: unbounded sibilant harmony

a.	/k-su-ʃojin/	k**ʃ**u**ʃ**ojin	'I darken it'
b.	/s-api-tʃʰo-us/	**s**api**tsʰ**olu**s**	'he has a stroke of good luck'
c.	/s-api-tʃʰo-us-waʃ/	**ʃ**api**tʃ**ʰolu**ʃ**wa**ʃ**	'he had a stroke of good luck'
d.	/k-su-k'ili-mekeken-ʃ/	k**ʃ**uk'ilimekeke**tʃ**	'I straighten myself up'

[3]Evidence for this includes Applegate's proposed relative ordering of the two phonological rules, the inclusion of both [st] and [sn] in his set of permitted word-medial consonant clusters, several data points included in the text, and an explicit statement that palatalization is reversed by the subsequent process of sibilant harmony (Applegate, 1972, p.120). It is also apparent that the misinterpretation of Applegate's description arises from a small list of exceptional words in which the local dissimilation occurs despite the presence of another [s] later in the word. I extend thanks to Jeff Heinz and Bill Idsardi (for further support for the description of the pattern provided in this section, see Heinz and Idsardi, 2010) for drawing my attention to the correct generalization of the pattern and suggesting that I look more closely at the data as presented in Applegate (1972).

The data in (34a)-(34c) demonstrate the local restriction against *[st, sn, sl], in which the prefix /s-/ surfaces as [ʃ] when it immediately precedes an alveolar consonant [t, n, l]. (34d) shows that the resulting [ʃ] may also serve as a trigger for sibilant harmony.

(34) Samala: /s/→[ʃ] when preceding (adjacent) [t, n, l]

a.	/s-tepuʔ/	**ʃt**epuʔ	'he gambles'
b.	/s-niʔ/	**ʃn**iʔ	'his neck'
c.	/s-lok'in/	**ʃl**ok'in	'he cuts it'
d.	/s-is-tɨʔ/	ʃi**ʃt**ɨʔ	'he finds it'

Finally, (35) shows that when both patterns cannot be upheld simultaneously, long-distance agreement is given priority over local disagreement. In (35a), an underlying /ʃ/ surfaces as [s] even though it immediately precedes a [t], because it is later followed by a non-adjacent [s] in the suffix. In (35b), the prefix /s-/ surfaces faithfully despite the following adjacent [n], in order to satisfy the requirement for agreement among sibilants.

(35) Samala: long-distance agreement overrides local disagreement

a.	/s-iʃ-tiʃi-jep-us/	sistisijepus	'they (dual) show him'
b.	/s-net-us/	snetus	'he does it to him'

The phonotactics of Samala present a challenge similar to the case of Tamashek Tuareg above (Section 5.3.2.1), in that it is not possible to capture the overall pattern with single TSL_2 grammar. Since both [st] and [sn] are observed in a string-adjacent context, they must be permitted as 2-factors on a tier that includes all segments (even though they are only permitted when a [+ant] segment such as [s] follows them later in the string). However, since [st] and [sn] are allowed to occur in such contexts, then a TSL_2 grammar would have no means of banning *[st] and *[sn] when there is no subsequent [s] in the string.

The tableaux in (36) and (37) demonstrate that two markedness constraints that are themselves TSL_2 grammars can be combined in an OT fashion to arrive at the correct pattern, when the higher ranked constraint operates on the tier of sibilants rather than the tier that includes all segments.

(36) Local disagreement in Samala

	$\left\langle \begin{array}{c} T = \{\text{s}, \text{ʃ}\} \\ R = \{\text{*sʃ, *ʃs}\} \end{array} \right\rangle$	$\left\langle \begin{array}{c} T = \Sigma \\ R = \{\text{*st, *sn, *sl}\} \end{array} \right\rangle$
a. ☞ ʃniʔ		
b. sniʔ		$*!$

(37) Non-adjacent agreement in Samala

	$\left\langle \begin{array}{c} T = \{\text{s}, \text{ʃ}\} \\ R = \{\text{*sʃ, *ʃs}\} \end{array} \right\rangle$	$\left\langle \begin{array}{c} T = \Sigma \\ R = \{\text{*st, *sn, *sl}\} \end{array} \right\rangle$
a. ʃnetus	$*!$	
b. ☞ snetus		$*$

5.3.3 TSL_2 Constraints in Phonological Theory

The patterns found in Tamashek Tuareg and Samala are both examples of languages that cannot be accounted for with a single TSL_2 grammar, nor with a conjunction of multiple TSL_2 grammars, since two different tiers permit/prohibit conflicting sets of 2-factors. Moreover, in the case of Tamashek Tuareg, these patterns co-exist with an *additional* TSL_2 pattern of sibilant harmony (see Section 5.3.1.1). As discussed in Section 5.3.2.1, ordered rules and ranked constraints have long been used in the phonological analysis of such patterns, and I have presented a preliminary illustration of how to do so using constraints that are defined as individual, violable and ranked TSL_2 grammars. A violation is assigned if the candidate does not belong to the stringset extension of the grammar, and the constraints can be ranked in any order. Intuitively, this family of constraints (i.e. those defined by TSL_2 grammars) does not differ greatly from the types of output well-formedness constraints that have been proposed in the literature as drivers of harmony and dissimilation, such as AGREE[F] or *XY for harmony and OCP[F] or *XX for dissimilation (see, e.g., Suzuki, 1998; Baković, 2000; Pulleyblank, 2002). From that perspective, the main contribution of the present work, aside from arguing that constraints of this kind are indeed necessary, is that it provides an extended definition of the constraint family that is computationally grounded. It is also of note that each of the patterns presented in Section 5.3.1 as a conjunction of TSL_2 languages can be generated in the same way, by ranking each of the conjoined grammars in any order (since they do not conflict, there will never be any evidence for determining the relative ranking of the two constraints).

While it would be interesting to explore whether or not the typology is skewed with respect to a particular ranking of two such constraints (e.g. the local constraints being ranked above non-local constraints), the literature does not provide enough examples to allow for such an assessment. This is in part due to the fact that in order for a language to exemplify two phonotactic patterns operating on two different tiers, the language must contain at least one long-distance dependency in the first place (e.g. non-adjacent labial dissimilation in Tamashek Tuareg, or sibilant harmony in Samala), and the sparsity of relevant patterns is an inevitable result of having limited the scope of my dissertation research to long-distance interactions between consonants in particular, which are relatively uncommon to begin with. Moreover, both of the examples presented above involve one constraint whose TSL_2 grammar operates on a tier that includes all segments.

To my knowledge, there is only one potential case that involves two conflicting dependencies between consonants, both of which apply in non-adjacent contexts. In Sundanese (see Cohn, 1992; Bennett, 2013, 2015), there is a pattern of (unbounded) liquid *dissimilation*, evidenced when the infix /-ar-/ surfaces as [-al-] when it precedes another [r] elsewhere in the word (e.g. [ŋab-a**r**-edo**l**] 'pull in' vs. [ŋ-a**l**-umba**r**a] 'go abroad'; Cohn, 1992, p. 206). However, another pattern of (transvocalic) liquid *harmony* overrides the dissimilation in some cases, resulting in words that include grammatical sequences of [rVr] or [lVl] (e.g. [**r**-a**r**-ɨwat] 'startled', [**l**-a**l**-ɨtik] 'little'; Cohn, 1992, p. 206). Unfortunately, the full pattern of Sundanese /-ar-/ infixation is not suitable for the present investigation, as it involves a number of descriptively complex conditioning factors that the TSL_2 account is not yet equipped to deal with (e.g. sensitivity to stem-initial vs. non-stem-initial position, root vs. affix affiliation, and onset vs. coda status; for further details, see Cohn, 1992; Bennett, 2015).

As an alternative source of empirical evidence that a Sundanese-like pattern should be included in the set of possible languages, it will be useful, in future research, to conduct an additional artificial language learning experiment. In particular, two new training conditions could be constructed (e.g. labelled M-Harm-S-Diss and M-Diss-S-Harm). For the M-Harm-S-Diss group, half of their training stems would contain liquids in the Medium-range context (CV<u>L</u>VCV) and exhibited a pattern of liquid harmony when suffixes [-li] or [-ɹu] were attached. The remaining half of stems, of the Short-range variety (CVCV<u>L</u>V), would instead exhibit a pattern of liquid dissimilation triggered by the suffixes. The structure of the training phase would be similar for subjects in the M-Diss-S-Harm group, but they would instead be exposed to liquid dissimilation at Medium-range and liquid harmony at Short-range. (Note that the patterns corresponding to the training phases of the proposed M-Harm-S-Diss and M-Diss-S-Harm groups can both be generated with TSL_2 constraints, but that only the latter can be derived within the factorial typology of ABC constraints.)

Even though the overall patterns cannot be described as individual TSL_2 languages, subjects are predicted to be able to learn the appropriate patterns. For example, the pattern corresponding to the training data of an M-Harm-S-Diss group could be generated with the use of TSL_2 constraints, as shown in (38) and (39).

(38) M-Harm-S-Diss target grammar: Medium-range harmony

Medium-range	$\left\langle \begin{array}{c} T = \{\text{C, l, ɹ}\} \\ R = \{\text{*lɹ, *ɹl}\} \end{array} \right\rangle$	$\left\langle \begin{array}{c} T = \{\text{l, ɹ}\} \\ R = \{\text{*ll, *ɹɹ}\} \end{array} \right\rangle$
a. ☞ CVlVCV-lV		
b. CVɹVCV-lV		$*!$

(39) M-Harm-S-Diss target grammar: Short-range dissimilation

Short-range	$\left\langle \begin{array}{c} T = \{\text{C, l, ɹ}\} \\ R = \{\text{*lɹ, *ɹl}\} \end{array} \right\rangle$	$\left\langle \begin{array}{c} T = \{\text{l, ɹ}\} \\ R = \{\text{*ll, *ɹɹ}\} \end{array} \right\rangle$
a. CVCVlV-lV	$*!$	
b. ☞ CVCVɹV-lV		$*$

The prediction that such patterns should be learnable can be tested with the same stimuli and procedures that were used in Experiments 1 through 4, and plans are currently underway to do so as a follow-up to this dissertation research.

5.3.4 Conclusion: The TSL_2 Region Is Not Too Small

The empirical evidence presented in this section suggests that the TSL_2 region of the subregular hierarchy does not offer enough complexity to account for certain patterns observed in natural language. However, each of the problematic patterns is a combination of multiple TSL_2 patterns that co-exist in a single language, which require different tiers to be specified for certain sets of 2-factor restrictions. In many cases—when there is no conflict between the two patterns—a single phonotactic grammar can be achieved with a simple conjunction of multiple TSL_2 grammars (see Section 5.3.1). There are a few additional cases in which this cannot be achieved, namely when two phonotactic generalizations are in direct conflict with one another and neither can be satisfied without violating the other (Section 5.3.2). Although there do not seem to be many such languages attested cross-linguistically, pilot results from an extension of Experiments 1 through 4 suggest that human learners are indeed able to learn such patterns in the lab. To account for this, I propose that individual TSL_2 grammars can be thought of as a family of constraints that can be integrated into various constraint-based frameworks. In particular, constraints of this type are attractive because they are similar to many other constraint families that have been proposed in literature (see Section 5.3.3), but are not subject to the same theoretical restrictions. This relative flexibility of TSL_2 constraints is precisely what allows them to straightforwardly account for descriptively complex patterns, such as long-distance dependencies with blocking by a relatively arbitrary set of intervening consonants. I believe that this, along with their computational properties of learnability, makes the TSL_2 definition of (long-distance co-occurrence) markedness constraints an attractive alternative to other constraint families that have been proposed in the literature as a means of deriving consonant harmony and long-distance dissimilation. I note, however, that the 2TSLIA (Jardine and Heinz, 2015) is only designed to discover *surface-true* patterns that belong to the TSL_2 region. Since the types of phonotactic systems that require constraint-like interactions are precisely those in which one pattern overrides another, rendering one of the patterns non-surface-true, further research must be done to determine what conditions are necessary in order to achieve efficient learning for the types of grammars that were considered in this section.

5.4 Summary and Conclusions

The purpose of this chapter was to support the argument for the TSL_2 approach to characterizing long-distance phonotactics and the human learner's hypothesis space for such patterns by asking three questions about the class of languages. I first argued in Section 5.1 that the TSL_2 region is not too big, as there exist certain patterns whose TSL_2 grammar requires specification of a relatively arbitrary set of segments as the tier. Furthermore, there is experimental evidence from Koo and Oh (2013) that human learners are capable of detecting an arbitrary set of dependencies on an arbitrarily defined tier, which supports the idea that they should indeed be included in the hypothesis space. However, the relative descriptive complexity of TSL_2 stringsets (as compared to, for example, the Strictly 3-Local or Strictly 2-Piecewise classes of formal languages; see Section 4.1) motivates the question of whether or not they are even computationally learnable. Section 5.2 provided a summary and an example implementation of the Tier-Based Strictly 2-Local Inference Algorithm (2TSLIA) proposed by Jardine and Heinz (2015), who prove that TSL_2 grammars are efficiently learnable. Finally, Section 5.3 presented several patterns that cannot be characterized as members of the TSL_2 class of stringsets. However, each pattern is made up of interacting dependencies that may override each other, and I have demonstrated how we can use TSL_2 grammars to define phonological markedness constraints that can be integrated into more familiar constraint-based frameworks and I argued that doing so offers a number of advantages from a computational perspective.

Chapter 6

Summary and Conclusions

6.1 Empirical Findings

This dissertation has established a set of empirical results that any theory of long-distance consonant phonotactics needs to be able to account for. With respect to locality relations, Table 6.1 summarizes the range of attested patterns.

Table 6.1: Typology of locality relations in patterns of consonant harmony (Harm) and long-distance consonant dissimilation (Diss). Note that the label '>transvocalic' refers to beyond-transvocalic locality.

Type	Locality	CvC	CvcvC	CvcvcvC	Attested?
	unbounded	+	+	+	✓
Harm	transvocalic	+	–	–	✓
	>transvocalic	–	+	+	✗
	unbounded	+	+	+	✓
Diss	transvocalic	+	–	–	✓
	>transvocalic	–	+	+	✗[1]

[1] Note that I classify beyond-transvocalic dissimilation as an unattested pattern, despite the potential empirical support from the case in Sundanese (Cohn, 1992; Bennett, 2013, 2015)—a complex pattern that I argue is better interpreted as a local (transvocalic) requirement for liquid harmony that overrides a more general restriction against the co-occurrence of identical liquids (see Section 5.3.3).

The typology in Table 6.1 raises the question of why there exists a robust dichotomy between unbounded and transvocalic dependencies, and why other logically possible patterns remain categorically unattested, such as a co-occurrence restriction that is enforced only in beyond-transvocalic contexts. In the above chapters, I have pursued in depth the idea that certain phonotactic patterns are unattested because there is no grammar within the human learner's hypothesis space that could generate that pattern. In other words, the learner is equipped with an inductive learning bias that renders certain patterns synchronically and diachronically inaccessible.

To further investigate these issues, I conducted a series of artificial language learning experiments, in which participants were tasked with learning a dependency between two liquid consonants from various permutations of training stimuli that resulted in nine training conditions (summarized in Table 6.2).

Table 6.2: Summary of training conditions for Experiments 1 through 4. Coloured cells, marked with '+' , indicate that evidence of a dependency that holds at that distance was presented in the training phase. Grey cells, marked with '–', indicate that the training stimuli included liquids at that distance, but that they always stayed faithful, and this did not conform to any systematic harmony or dissimilation pattern. White cells with '?' are contexts for which no exposure to liquids was provided during the training phase.

Exp.	Group	…Lv-Lv	…Lvcv-Lv	Lvcvcv-Lv
1	M-Harm	?	+	?
	S-Harm	+	?	?
2	M-Diss	?	+	?
	S-Diss	+	?	?
3	M-Harm-S-Faith	–	+	?
	S-Harm-M-Faith	+	–	?
4	M-Diss-S-Faith	–	+	?
	S-Diss-M-Faith	+	–	?
1-4	Control	?	?	?

I argued that the results of Experiments 1 through 4 provide relatively strong evidence in support of the hypothesis that the typology of long-distance dependencies is a reflection of a human learning bias.

In Experiments 1 and 2, human learners were presented with training data that did not offer complete information about the exact nature of the pattern, and the general results were the same whether the target pattern was liquid harmony (Experiment 1) or liquid dissimilation (Experiment 2). Recall that the participants in the M-Harm (Exp. 1) and M-Diss (Exp. 2) training conditions were exposed to pairs of liquids that were separated by a Medium-range distance (cvLvcv-Lv), but did not encounter any data on Short-range (cvcvLv-Lv) or Long-range (Lvcvcv-Lv) distances. Only one attested (and by hypothesis possible) type of locality is compatible with the evidence presented in their training phase (namely, unbounded). Indeed, even though none of the participants had any prior experience with such a pattern, they tended to internalize it with unbounded locality, applying the dependency to all three distance in the testing phase (as opposed to, e.g., learning a dependency that holds specifically between a liquid in the second syllable of the stem and a liquid in the suffix). For participants in the S-Harm (Exp. 1) and S-Diss (Exp. 2) groups, the training data included evidence of a phonotactic dependency between liquids at Short-range distances, but they received no exposure to pairs of liquids in Medium- or Long-range contexts. Such a pattern is, in principle, compatible with either of the attested transvocalic or unbounded locality variants, but learners tended to interpret the pattern as strictly-transvocalic, and only a small effect (if any) was observed at the group level for generalizing the target pattern to greater distances. Experiment 1 thus replicates the findings of previous experiments looking at the learning of locality relations in patterns of sibilant harmony (Finley, 2011, 2012; McMullin and Hansson, 2014, Experiment 1), and furthermore extends them to a different class of segments (i.e. liquids rather than sibilants). Likewise, the results of Experiment 2 suggest that humans learn and generalize phonotactic dependencies in the same way, whether the nature of the interaction is assimilatory or dissimilatory.

In Experiments 3 and 4, the training phase provided participants with information about the behaviour of liquids in both in both Short-range and Medium-range contexts. For each of the experimental groups, the training phase exhibited a suffix-triggered pattern of liquid harmony at one of the two distances, but, at the other distance, there was no restriction on the co-occurrence of liquids. Note that the resulting patterns shown to the S-Harm-M-Faith (Exp. 3) and the S-Diss-M-Faith (Exp. 4) groups are not in conflict with the typology, as such patterns are compatible with the well-attested transvocalic variants of consonant harmony and dissimilation, respectively. As expected, this did not impede learning, and the subjects in both groups tended to apply the pattern exactly as it was evidenced in training. By contrast, the patterns exhibited in the training phases for the M-Harm-S-Faith (Exp. 3) and M-Diss-S-Faith (Exp. 4) groups, which exhibited phonotactic restrictions that held at Medium-range (triggering liquid alternations), but not at Short-range (overt evidence of faithful non-alternation), are not compatible with any type of attested pattern in terms of locality. Specifically, these beyond-transvocalic patterns violate the observed typological universal that if a language enforces a particular phonotactic restriction in Medium-range contexts, then the same restriction applies at Short-range. (The contrapositive is also true—no restriction at Short-range implies no Medium-range restriction.) Not surprisingly, such patterns prove to be extremely difficult to for humans to learn in an artificial language learning task, and very few individual participants in these training conditions successfully learned that a dependency held in Medium-range contexts. Furthermore, of those that did, the majority over-generalized, applying the same phonotactic pattern to pairs of liquids in Short-range contexts, in spite of the overt evidence in the training data that transvocalic pairs of liquids should remain faithful. In sum, I interpret the results of Experiments 1 through 4 as evidence for the connection between typology and human learning bias.

6.2 Assessing Theoretical Approaches

In light of the above empirical data, this dissertation has assessed the predictions of two distinct theoretical frameworks in terms of whether or not they offer a satisfactory account of the boundary between phonotactic patterns that are human-learnable and those that are not.

Agreement by Correspondence Within Optimality Theory (Prince and Smolensky, 2004), I focused primarily on the Agreement by Correspondence framework (Walker, 2000a,c; Hansson, 2001, 2010a; Rose and Walker, 2004), which has seen relative success in accounting for the typology of consonant harmony, and which has recently been extended as a comprehensive analysis of long-distance consonant dissimilation (Bennett, 2013). From this perspective, the boundary between a possible and impossible language can be be defined directly in terms of the factorial typology of the universal constraint set assumed by ABC. Two main strategies have been proposed as a means of dealing with locality in ABC. Interestingly, although each relies on some version of a locality-based constraint, they result in different sets of patterns that can or cannot be generated. The first approach is to define a CC·Limiter constraint that penalizes correspondence outside of the relevant window, such as PROXIMITY (defined by Rose and Walker, 2004, in terms of syllable-adjacency; redefined in Chapter 2 with respect to transvocalic contexts), or CC-CVC (based on CC·SYLLADJ; Bennett, 2013). The second strategy is to define a CORR[αF] constraint that enforces correspondence only within the bounded range (e.g. CORR-CVC[αF]). Originally advocated by Hansson (2001, 2010a) as an alternative to constraints like PROXIMITY or CC-CVC, Bennett (2013) argues that they are necessary *in addition* to a locality-based CC·Limiter.

Table 6.3 summarizes the types of patterns that can be generated within a factorial typology of five ABC constraints: four basic constraints (including CORR[αF], CC-IDENT[G], IO-IDENT[F], IO-IDENT[G]), and one of PROXIMITY, CC-CVC, or CORR-CVC[αF]. In this schematic overview of the predictions, note that each pattern of harmony would enforce agreement for [G] among those consonants with a particular set of shared features (one of which is [F]), whereas the patterns of dissimilation would require disagreement for [F].[2]

With respect to consonant harmony, the ABC model makes all of the correct predictions no matter which of the three proposed locality-based constraints is used, PROXIMITY, CC-CVC, or CORR-CVC[αF]. This is no doubt a reflection of the fact that the surface-correspondence approach was developed specifically as an analysis of consonant harmony and the unique cross-linguistic properties exhibited by such

[2]Recall that, all else equal, IO-IDENT[G] $\gg$ IO-IDENT[F] favours a pattern of consonant harmony, and IO-IDENT[F] $\gg$ IO-IDENT[G] favours dissimilation.

Table 6.3: Types of phonotactic patterns and locality relations that can (✓) or cannot (✗) be generated within the factorial typology of ABC, using different locality-based constraints. Green and red icons indicate whether the prediction matches the typology or not, respectively.

Type	Locality	Attested?	Constraint version		
			PROXIMITY	CC-CVC	CORR-CVC
Harm	unbounded	✓	✓	✓	✓
	transvocalic	✓	✓	✓	✓
	>transvocalic	✗	✗	✗	✗
Diss	unbounded	✓	✓	✓	✓
	transvocalic	✓	✗	✗	✓
	>transvocalic	✗	✓	✓	✗

patterns. However, PROXIMITY and CC-CVC, which penalize correspondence outside of transvocalic contexts, predict that beyond-transvocalic patterns of *dissimilation* should be attested, while the simple (and widely attested) transvocalic variant of dissimilation should not (cf. the "Mismatch Prediction" about the typologies of consonant harmony vs. dissimilation; Bennett, 2013; discussed in Section 3.1.2). Moreover, Section 3.1.1 demonstrated that the global evaluation of PROXIMITY leads to pathological patterns, such as dependencies that are sensitive to the count (even vs. odd parity) of potential correspondents. CC-CVC seems to provide some improvement, since it is evaluated only for those pairs of surface-corresponding consonants that are "local", in the sense of not being separated by a member of the same correspondence class. Thus, in a sequence $\ldots C_x \ldots C_x \ldots C_x \ldots$ the $C_1 \leftrightarrow C_2$ and $C_2 \leftrightarrow C_3$ correspondent pairs are subject to CC-CVC but not the $C_1 \leftrightarrow C_3$ pair, much as though the correspondents are being treated as a tier (ordered subsequences of the output string) rather than as the unordered sets (equivalence classes) defined by the formal correspondence relation. The alternative strategy of using CORR-CVC[αF] constraints, which enforce correspondence only in transvocalic contexts, seems to provide a much better fit to the empirical data. Notice, however, that this move indirectly incorporates the notion of a "consonant tier", in that the presence of any intervening consonant nullifies the demand for correspondence (and thereby the motivation for harmony or dissimilation) between the segments of interest.

It is notable that additional proposals for modifying the evaluation of ABC constraints also trend in the direction of treating collections of surface-correspondents as if they were tiers. Hansson (2007) suggests, for example, that CC-IDENT constraints can be evaluated only for segment pairs that are adjacent in the "correspondence chain". The motivation, again, is to avoid pathological predictions of the factorial typology, which includes harmony systems where a "majority rule" (Lombardi, 1999; Baković, 2000) determines the directionality of assimilation, or even the status of an intervening segment as opaque vs. transparent to the harmony, on a word-by-word basis. I note, however, that in spite of all of the proposed (tier-like) changes that seem to improve the predictions, the inherently complex ABC machinery still permits certain pathologies. In particular, I draw attention to cases of "agreement by proxy" (Hansson, 2014; discussed at length in Section 3.1.1.1), in which two relatively dissimilar segments (e.g. [s] and [g]) are forced to be in surface-correspondence (and therefore to agree in some way) if and only if each one is sufficiently similar to a third consonant (e.g. [x] occurring somewhere else in the word). Such patterns are expected to arise due to the transitivity of the surface-correspondence relation, whereby in a sequence $\ldots C_x \ldots C_x \ldots C_x \ldots$, if $C_1 \leftrightarrow C_3$ and $C_2 \leftrightarrow C_3$ pairs are in surface correspondence, then so is the $C_1 \leftrightarrow C_2$ pair. It seems that this prediction cannot be avoided unless the idea of an all-encompassing correspondence relation is abandoned altogether.

Based on the above evidence, this dissertation argued that the factorial typology of ABC constraints does not (and cannot) provide an accurate characterization of the set of possible patterns, and therefore that we need to pursue a different theoretical account of the typology and learnability of long-distance phonotactics.

Subregular Stringsets As an alternative approach, I investigated potential solutions within the framework of formal language theory, in which the phonotactics of a language can be thought of as a systematic distinction between the set of grammatical vs. ungrammatical words (where words are strings of segments). One of the conceptual draws of this approach is that the properties of phonotactic dependencies can be investigated computationally, and the proposed space of possible patterns can be expressed as a well-defined class of formal languages. More specifically, this dissertation explored long-distance phonotactic patterns in terms of where they are situated within the subregular hierarchy (see, e.g., McNaughton and Papert, 1971; Rogers et al., 2010; Heinz et al., 2011; Rogers and Pullum, 2011), focusing primarily on three such classes: the Strictly 3-Local (SL_3), Strictly 2-Piecewise (SP_2), and Tier-based Strictly 2-Local (TSL_2) languages. A summary of the types of patterns that are contained within each of these regions (as well as the region defined by the union of the SL_3 and SP_2 classes) is provided in Table 6.4.

Table 6.4: Types of phonotactic patterns and locality relations that can (✓) or cannot (✗) be generated as members of different subregular classes of formal languages (stringsets). Green and red icons indicate whether the prediction matches the typology or not, respectively.

			Subregular class			
Type	Locality	Attested?	SL_3	SP_2	$SL_3 \cup SP_2$	TSL_2
Harm	unbounded	✓	✗	✓	✓	✓
	transvocalic	✓	✓	✗	✓	✓
	>transvocalic	✗	✗	✗	✗	✗
Diss	unbounded	✓	✗	✓	✓	✓
	transvocalic	✓	✓	✗	✓	✓
	>transvocalic	✗	✗	✗	✗	✗

While neither the SL_3 nor the SP_2 class contains all of the desired patterns, one possibility is that long-distance dependencies can be either SL_3 or SP_2 (Heinz, 2010). More specifically, McMullin and Hansson (2014) argue that the human phonotactic learner consists of (at least) two modules for learning long-distance consonantal phonotactics: an *n*-gram learner for acquiring transvocalic patterns (as SL_3 languages that ban certain $*\underline{\text{C}}\text{v}\underline{\text{C}}$ trigrams, or 3-factors), as well as a precedence learner (Heinz, 2010) that is responsible for detecting unbounded dependencies (as SP_2 languages with restrictions on certain $\underline{\text{C}}\ldots\underline{\text{C}}$ subsequences). The resulting characterization of the learner's hypothesis space is the union of these two classes of formal languages ($SL_3 \cup SP_2$), which, as shown in Table 6.4, reflects the range of locality relations that are observed in patterns of consonant harmony and long-distance consonant dissimilation.

The TSL_2 class of formal languages (Heinz et al., 2011) likewise provides a sufficient account of the empirical findings with respect to locality relations. Recall that TSL_2 languages are defined by a tier T (a subset of the segment inventory Σ), and a set of 2-factors that are prohibited in tier-adjacent contexts (labelled R; or alternatively, a set S of 2-factors that are permitted on the tier). From this perspective, the difference between the transvocalic and unbounded variants of long-distance dependencies is whether the grammaticality of each word is assessed on the tier of consonants (where only those consonant pairs in a $\underline{C}(v)\underline{C}$ relationship can potentially violate the phonotactics), or a tier comprised only of potential triggers or targets (e.g. the liquid tier, the sibilant tier, etc.).

As both of the proposed regions (i.e. the TSL_2 class and the union of the SL_3 and SP_2 classes) seem to offer a good definition of the boundary between possible and impossible patterns with respect to locality relations, I compared additional types of patterns contained within each region and argued that there is more evidence that supports the TSL_2 approach. Specifically, I argued that treating the two elements of a dependency as adjacent segments on a tier allows for a unified account of both locality and blocking in long-distance phonotactics. The difference between whether a long-distance consonant interaction can or cannot be blocked by a specific intervening segment is simply whether or not that segment is a member of the relevant tier. As an example of this, Table 6.5 demonstrates that specifying different sets of segments as members of the tier results in exactly the types of patterns that are attested cross-linguistically.

Table 6.5: Illustration of how the grammaticality of words with sibilant subsequences varies as a result of modifying the contents of the tier specified by a TSL_2 grammar, with a segment inventory $\Sigma = \{s, ʃ, t, p, a\}$.

Sibilant harmony	Tier	Word (grammatical or not?)				
($R = \{$*sʃ, *ʃs$\}$)		sapas	sapaʃ	sataʃ	saʃ	asʃa
unbounded	{s, ʃ}	✓	✗	✗	✗	✗
blocking	{s, ʃ, t}	✓	✓	✗	✗	✗
transvocalic	{s, ʃ, t, p}	✓	✓	✓	✗	✗
(direct-adjacency)	{s, ʃ, t, p, a}	✓	✓	✓	✓	✗

Although long-distance dependencies with blocking are relatively rare across the world's languages, they are indeed attested for both assimilatory and dissimilatory interactions between consonants. Since these patterns cannot be captured within the $SL_3 \cup SP_2$ region, I concluded that the class of TSL_2 formal languages provides a close approximation of the range of patterns that are supported empirically.

Finally, it is important to note that formal language theory and phonological theory are not inherently incompatible, and I argue that the two approaches offer a mutual benefit. For example, many constraints that have been proposed in order to account for long-distance interactions are defined in ways that very closely resemble a TSL_2 grammar (where a violation is assigned if a particular candidate is not a member of the corresponding stringset). I suggest that stating them in formal terms allows us to better understand the range of predictions and to account for many patterns that are otherwise rather difficult to handle within the confines of phonological theory (e.g. segmental blocking effects, arbitrary tiers, etc.). Likewise, an individual TSL_2 grammar has no way of accounting for complex patterns that quite clearly arise due to the interaction of two phonotactic restrictions, in which one overrides the other when they cannot both be satisfied simultaneously. There is no clear path to dealing with this strictly in terms of formal language theory, but the integration of formal grammars with the well-studied notion of constraint rankings (or rule orderings) often allows for relatively simple solutions for analyzing otherwise complex phonotactic interactions. Finally, I point out that while Jardine and Heinz (2015) have recently proposed an algorithm (the 2TSLIA; see Section 5.2.1) that can provably and efficiently acquire a correct grammar for any individual TSL_2 pattern, further investigation into the learnability of multiple TSL_2 patterns (that are not necessarily surface-true) is needed before this strategy can be considered computationally tractable.

6.3 Outstanding Issues

As the scope of this dissertation was restricted to locality relations in long-distance consonantal phonotactics, there are a number of issues that need to be addressed. In this section, I briefly discuss four important areas of research that deserve attention in future work: segmental blocking, the role of phonological similarity, other types of non-adjacent dependencies, and the computational properties of input-output mappings.

First, there is a need for further empirical investigation of the learnability of blocking effects in patterns of consonant harmony and long-distance consonant dissimilation. Due to the cross-linguistic sparsity of such patterns, there is much to be gained from the use of an artificial language learning paradigm to study issues related to their learnability. However, to my knowledge, no such study has been conducted, and there are a number of questions that need to be addressed. For example (among many others): Can long-distance consonant interactions that are blocked by certain intervening segments be learned in the laboratory? If so, what conditions are necessary to achieve learning? How do learners generalize from limited information in the training phase? Can any segment be a blocker? Do the predictions align with the set of attested patterns (few as they may be), and the predictions of the TSL_2 approach?

With respect to phonological similarity and natural classes, I have argued that an advantage of TSL_2 grammars is the ability to specify a tier that contains any (potentially arbitrary) set of segments. However, it is clear that the majority of attested patterns can be characterized with a TSL_2 language whose tier is indeed a set of segments that form a natural class (e.g., the sibilants, the liquids, the voiced obstruents, etc.). At present, it remains unclear how features (or other representational structure) are best treated in terms of formal language theory. However, future advances in this area may shed light on the relationship between how the Tier-based Strictly 2-Local Inference Algorithm (2TSLIA; Jardine and Heinz, 2015) learns phonotactic patterns (i.e. with no preference for 'natural' tiers), and the biases exhibited by human learners in artificial language learning studies that consider various types of feature interactions (see, e.g., Wilson, 2003; Moreton, 2008, 2012; Koo and Oh, 2013).

As a future source of data, the TSL_2 approach can be extended beyond the analysis of long-distance interactions between consonants. For example, Jardine (2015) shows that the 2TSLIA (Jardine and Heinz, 2015, see Section 5.2.1 above) can be used to learn a TSL_2 grammar for the phonotactics of vowels in Finnish (using data from Goldsmith and Riggle, 2012), which exhibits several interesting properties of locality, transparency, and blocking in vowel harmony. Patterns of vowel harmony are widespread cross-linguistically, and there also exist certain interactions that hold between consonants and vowels (e.g. nasal harmony). Patterns like these may therefore provide an excellent empirical testing ground for the predictions of the TSL_2 approach (for details and previous analyses of vowel harmony and vowel-consonant harmony, see, e.g., van der Hulst and van de Weijer, 1995; Walker, 2000b; Archangeli and Pulleyblank, 2007; Finley, 2008; Nevins, 2010).

Finally, the treatment of long-distance dependencies as members of the Tier-based Strictly 2-Local class of formal languages is rather limited, in that it only offers an account of the phonotactic restrictions on surface forms. However, recent research into the formal characterization of input-output mappings has focused on establishing a hierarchy of well-defined classes of subregular *relations* (as opposed to stringsets), and investigating their associated computational properties, such as relative complexity and learnability (see, e.g., Chandlee, 2014; Chandlee and Jardine, 2014; Chandlee et al., 2014; Jardine et al., 2014; Payne, 2014). Although there is still much work to be done in this area, present results from the literature suggest that this approach to phonological mappings can provide an attractive alternative to constraint-based frameworks, and can be used to assess the computational implications of particular constraint sets.

6.4 Conclusion

With an increasing amount of support for the hypothesis that typological distributions of phonological patterns are shaped, in part, by human learning biases, it is important to understand the range of patterns that any theoretical model predicts to be possible and human-learnable. With respect to phonotactic patterns that can be generated in a constraint-based framework, this is not necessarily an easy task, since complex interactions of seemingly unrelated constraints can inadvertently over-generate, resulting in a number of pathologies. I argue that pursuing questions of pattern complexity and learnability within formal language theory can offer us a ‘computational grounding’ of phonology that may help to rectify certain problematic predictions, especially with respect to the structural properties of phonological patterns (e.g. locality and opacity), in the same way that ‘phonetic grounding' is an attempt to confine predictions in terms of substance (e.g. perceptual similarity). While there is much work to be done before we fully understand the computational properties of phonological patterns in natural language, each step of the pursuit can both advance our knowledge of the limits on human learnability and bring us closer to a cohesive explanation of phonological typology.

Bibliography

Alderete, J. and Frisch, S. A. (2007). Dissimilation in the grammar and the lexicon. In de Lacy, P., editor, *The Cambridge Handbook of Phonology*, pages 379–398. Cambridge University Press, Cambridge. → pages 8

Applegate, R. B. (1972). *Ineseño Chumash grammar*. Doctoral dissertation, University of California, Berkeley. → pages ii, 1, 140

Archangeli, D. and Pulleyblank, D. (1994). *Grounded phonology*. MIT Press, Cambridge, MA. → pages 13

Archangeli, D. and Pulleyblank, D. (2007). Harmony. In de Lacy, P., editor, *The Cambridge Handbook of Phonology*, pages 353–378. Cambridge University Press, Cambridge. → pages 158

Baković, E. (2000). *Harmony, dominance and control*. Doctoral dissertation, Rutgers University. → pages 142, 153

Bates, D., Maechler, M., and Bolker, B. (2014). *lme4: Linear mixed-effects models using S4 classes*. R package version 1.1-6. → pages 32

Becker, M., Nevins, A., and Levine, J. (2012). Asymmetries in generalizing alternations to and from initial syllables. *Language*, 88(2):231–268. → pages 69, 75

Bennett, W. (2013). *Dissimilation, consonant harmony, and surface correspondence*. Doctoral dissertation, Rutgers University. → pages ii, 6, 7, 8, 9, 10, 12, 14, 18, 21, 39, 40, 48, 49, 53, 54, 55, 56, 57, 59, 61, 62, 78, 79, 85, 131, 143, 147, 151, 152

Bennett, W. (2015). Assimilation, dissimilation, and surface correspondence in Sundanese. *Natural Language and Linguistic Theory*, 33(2):371–415. → pages 14, 49, 59, 62, 78, 143, 147

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press, Cambridge. → pages 5, 17, 113

Boersma, P. (1997). How we learn variation, optionality, and probability. In *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, volume 21, pages 43–58. → pages 13, 44

Boersma, P. and Hayes, B. (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry*, 32(1):45–86. → pages 44

Chandlee, J. (2014). *Strictly Local phonological processes*. PhD thesis, University of Delaware. → pages 158

Chandlee, J., Eyraud, R., and Heinz, J. (2014). Learning Strictly Local subsequential functions. *Transactions of the Association for Computational Linguistics*, 2:491–503. → pages 158

Chandlee, J. and Jardine, A. (2014). Learning phonological mappings by learning strictly local functions. In Kingston, J., Moore-Cantwell, C., Pater, J., and Staubs, R., editors, *Proceedings of the 2013 Meeting on Phonology*, Washington, DC. Linguistic Society of America. → pages 158

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory*, 2(3):113–124. → pages 15, 79

Clements, G. N. (1980). *Vowel harmony in nonlinear generative phonology: an autosegmental model*. Indiana University Linguistics Club, Bloomington, IN. → pages 17, 113

Clements, G. N. (1985). The geometry of phonological features. *Phonology Yearbook*, 2:225–252. → pages 17, 113

Clements, G. N. and Hume, E. V. (1995). The internal organization of speech sounds. In Goldsmith, J. A., editor, *The handbook of phonological theory*, pages 245–306. Blackwell, Oxford. → pages 17, 113

Cohn, A. C. (1992). The consequences of dissimilation in Sundanese. *Phonology*, 9:199–220. → pages 49, 59, 62, 143, 147

Coupez, A. (1980). *Abrège de grammaire Rwanda*. Institut National de Recherche Scientifique, Butare. → pages 122

Creel, S. C., Newport, E. L., and Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 30(5):1119–1130. → pages 1, 9

Cser, A. (2010). The *alis / aris* allomorphy revisited. In Rainer, F., Dressler, W. U., Kastovsky, D., and Luschützky, H. C., editors, *Variation and change in morphology: Selected papers from the 13th International Morphology Meeting, Vienna, February 2008*, pages 33–51. John Benjamins, Amsterdam. → pages 12, 118, 119, 122

Culbertson, J. (2012). Typological universals as reflections of biased learning: Evidence from artificial language learning. *Linguistics and Language Compass*, 6:310–329. → pages 3

Culbertson, J., Smolensky, P., and Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, 122:306–329. → pages 3

Culy, C. (1985). The complexity of the vocabulary of Bambara. *Linguistics and Philosophy*, 8:345–351. → pages 15

de la Higuera, C. (1997). Characteristic sets for polynomial grammatical inference. *Machine Learning*, 27(2):125–138. → pages 124

Dolbey, A. and Hansson, G. Ó. (1999). The source of naturalness in synchronic phonology. In Billings, S. J., Boyle, J. P., and Griffith, A. M., editors, *Papers from the 35th meeting of the Chicago Linguistic Society*, volume 1, pages 59–69, Chicago, IL. Chicago Linguistic Society. → pages 24, 76

Dresher, E. (1999). Charting the learning path: cues to parameter setting. *Linguistic Inquiry*, 30:27–67. → pages 13

Dressler, W. (1971). An alleged case of nonchronological rule insertion. *Linguistic Inquiry*, 2:597–599. → pages 12, 118, 119

Eisner, J. (1997). What constraints should OT allow? Paper presented at the 71st Annual Meeting of the Linguistic Society of America, Chicago, January 1997. [ROA-204 (talk handout)]. → pages 13

Ellison, T. M. (1992). *The machine learning of phonological structure*. PhD thesis, University of Western Australia. → pages 14

Ellison, T. M. (1994). Phonological derivation in optimality theory. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING '94)*, volume 2, pages 1007–1013, Stroudsburg, PA. Association for Computational Linguistics. [ROA-75]. → pages 14

Elmedlaoui, M. (1995). *Aspects des représentations phonologiques dans certaines langues chamito-sémitiques*. PhD thesis, Université Mohammed V. → pages 12, 88, 131, 134, 135

Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., and Wong, P. C. (2015). The relationship between artificial and second language learning. *Cognitive Science*. → pages 6

Fallon, P. D. (1993). Liquid dissimilation in Georgian. In Kathol, A. and Bernstein, M., editors, *Proceedings of the 10th Eastern States Conference on Linguistics*, pages 105–116, Ithaca, NY. DMLL Publications. → pages 7, 85, 86

Finley, S. (2008). *Formal and cognitive restrictions on vowel harmony*. PhD thesis, Johns Hopkins University. → pages 2, 158

Finley, S. (2011). The privileged status of locality in consonant harmony. *Journal of Memory and Language*, 65:74–83. → pages 5, 11, 24, 25, 39, 46, 73, 149

Finley, S. (2012). Testing the limits of long-distance learning: learning beyond a three-segment window. *Cognitive Science*, 36:740–756. → pages 5, 11, 24, 25, 39, 46, 73, 75, 149

Finley, S. and Badecker, W. (2007). Towards a substantively biased theory of learning. *Berkeley Linguistics Society*, 33:142–154. → pages 3

Finley, S. and Badecker, W. (2009). Artificial language learning and feature-based generalization. *Journal of Memory and Language*, 61(3):423–437. → pages 5, 93

Finn, A. S. and Hudson Kam, C. L. (2008). The curse of knowledge: first language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, 108:477–499. → pages 6

Gafos, A. I. (1999). *The articulatory basis of locality in phonology*. Garland, New York. → pages 6, 25

Garcia, P., Vidal, E., and Oncina, J. (1990). Learning locally testable languages in the strict sense. In Arikawa, S., Goto, S., Ohsuga, S., and Yokomori, T., editors, *Algorithmic Learning Theory, First International Workshop, ALT '90*, pages 325–338. Springer/Ohmsha. → pages 84

Garrett, A. and Johnson, K. (2012). Phonetic bias in sound change. In Yu, A. C. L., editor, *Origins of sound change: Approaches to phonologization*. Oxford University Press, Oxford. → pages 5

Gebhart, A. L., Newport, E. L., and Aslin, R. N. (2009). Statistical learning of adjacent and non-adjacent dependencies among non-linguistic sounds. *Psychonomic Bulletin & Review*, 16:486–490. → pages 1, 9

Gold, E. M. (1967). Language identification in the limit. *Information and Control*, 10:447–474. → pages 84, 124, 137

Goldsmith, J. A. and Riggle, J. (2012). Information theoretic approaches to phonological structure: the case of Finnish vowel harmony. *Natural Language and Linguistic Theory*, 30:859–896. → pages 2, 157

Goldwater, S. and Johnson, M. (2003). Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the Workshop on Variation within Optimality Theory*, Department of Linguistics: Stockholm University. → pages 13, 44

Halle, M. and Vergnaud, J.-R. (1981). Harmony processes. In Klein, W. and Levelt, W., editors, *Crossing the boundaries in linguistics: Studies presented to Manfred Bierwisch*, pages 1–22. Reidel, Dordrecht. → pages 6

Hansson, G. Ó. (2001). *Theoretical and typological issues in consonant harmony*. Doctoral dissertation, University of California, Berkeley. → pages 6, 8, 11, 14, 39, 48, 61, 78, 79, 84, 151

Hansson, G. Ó. (2007). Blocking effects in agreement by correspondence. *Linguistic Inquiry*, 38(2):395–409. → pages 12, 88, 114, 153

Hansson, G. Ó. (2008). Diachronic explanations of sound patterns. *Language and Linguistics Compass*, 2(5):859–893. → pages 5

Hansson, G. Ó. (2010a). *Consonant harmony: long-distance interaction in phonology*. University of California Press, Berkeley, CA. → pages ii, 6, 7, 8, 10, 11, 14, 18, 21, 22, 25, 39, 48, 78, 79, 82, 114, 131, 134, 135, 140, 151

Hansson, G. Ó. (2010b). Long-distance voicing assimilation in Berber: spreading and/or agreement? In Heijl, M., editor, *Actes du Congrès de l'ACL 2010 / 2010 CLA Conference Proceedings*. Canadian Linguistic Association. → pages 12, 61, 88, 131, 134, 135

Hansson, G. Ó. (2014). (Dis)agreement by (non)correspondence: inspecting the foundations. Paper presented at ABC↔Conference, Berkeley. [Slides published in *2014 UC Berkeley Phonology Lab Annual Report: ABC↔Conference Archive*, 3–62. Online: http://linguistics.berkeley.edu/phonlab/annual_ report/documents/2014/annual_report_2014_ABCC.html]. → pages 49, 56, 153

Hayes, B. (1999). Phonetically driven phonology: the role of optimality theory and inductive grounding. In Darnell, M., Moravscik, E., Newmeyer, F. J., Noonan, M., and Wheatly, K., editors, *Functionalism and formalism in linguistics*, volume 1, pages 243–285. John Benjamins. → pages 13

Hayes, B. (2004). Phonological acquisition in Optimality Theory. In Kager, R., Pater, J., and Zonneveld, W., editors, *Fixing priorities: constraints in phonological acquisition*, pages 158–203. Cambridge University Press, Cambridge. → pages 44

Hayes, B. and Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39:379–440. → pages 2, 14

Hayward, R. J. (1982). Notes on the Koyra language. *Afrika und Übersee*, 65:211–268. → pages 10, 21, 22, 53, 82

Hayward, R. J. (1990). Notes on the Aari language. In Hayward, R. J., editor, *Omotic language studies*, pages 425–493. School of Oriental and African Studies, University of London, London. → pages 10, 21, 83

Heath, J. (2005). *A grammar of Tamashek (Tuareg of Mali)*. Mouton de Gruyter, Berlin. → pages 131, 132, 133, 134, 137

Heinz, J. (2007). *Inductive learning of phonotactic patterns*. Doctoral dissertation, University of California, Los Angeles. → pages 2, 80, 84

Heinz, J. (2009). On the role of locality in learning stress patterns. *Phonology*, 26:303–351. → pages 13

Heinz, J. (2010). Learning long-distance phonotactics. *Linguistic Inquiry*, 41(4):623–661. → pages 2, 15, 17, 79, 80, 81, 83, 84, 85, 87, 154

Heinz, J. and Idsardi, W. J. (2010). Learning opaque generalizations: the case of Samala. Unpublished ms. University of Delaware, University of Maryland. → pages 140

Heinz, J. and Idsardi, W. J. (2011). Sentence and word complexity. *Science*, 333:295–297. → pages 84

Heinz, J., Kasprzik, A., and Kötzing, T. (2012). Learning in the limit with lattice-structured hypothesis spaces. *Theoretical Computer Science*, 457:111–127. → pages 137

Heinz, J., Rawal, C., and Tanner, H. G. (2011). Tier-based strictly local constraints for phonology. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 58–64, Portland, OR. Association for Computational Linguistics. → pages iii, 2, 15, 17, 18, 79, 80, 85, 86, 87, 90, 111, 124, 131, 154, 155

Hudson Kam, C. L. and Newport, E. L. (2005). Regularizing unpredictable variation: the roles of adult and child learners in language variation and change. *Language Learning and Development*, 1:151–195. → pages 5

Hudson Kam, C. L. and Newport, E. L. (2009). Getting it right by getting it wrong: when learners change languages. *Cognitive Psychology*, 59:30–66. → pages 5

Hyman, L. M. (1995). Nasal consonant harmony at a distance: the case of Yaka. *Studies in African Linguistics*, 24:5–30. → pages 22

Jardine, A. (2015). Learning tiers for long-distance phonotactics. Unpublished ms. University of Delaware. [To appear in *Proceedings of the 6th Conference on Generative Approaches to Language Acquisition North America (GALANA 2015)*]. → pages 124, 157

Jardine, A., Chandlee, J., Eyraud, R., and Heinz, J. (2014). Very efficient learning of structured classes of subsequential functions from positive data. In Clark, A., Kanazawa, M., and Yoshinaka, R., editors, *Proceedings of the 12th International Conference on Grammatical Inference (ICGI 2014)*, volume 34 of *JMLR Workshop Proceedings*, pages 94–108. → pages 158

Jardine, A. and Heinz, J. (2015). Learning Tier-based Strictly 2-Local languages. Unpublished ms. University of Delaware. → pages 2, 17, 87, 122, 123, 124, 125, 126, 130, 131, 145, 146, 156, 157

Jensen, J. (1974). Variables in phonology. *Language*, 50:675–686. → pages 12, 118, 119

Jesney, K. and Tessier, A.-M. (2011). Biases in Harmonic Grammar: the road to restrictive learning. *Natural Language and Linguistic Theory*, 29:251–290. → pages 44

Johnson, C. D. (1972). *Formal aspects of phonological description*. Mouton, The Hague. → pages 14, 79

Jurgec, P. (2011). *Feature spreading 2.0: a unified theory of assimilation*. PhD thesis, University of Tromsø. → pages 12, 88, 123

Kaplan, R. M. and Kay, M. (1994). Regular models of phonological rule systems. *Computational Linguistics*, 20(3):331–378. → pages 14, 79

Kimenyi, A. (1979). *Studies in Kinyarwanda and Bantu phonology*. Linguistic Research, Carbondale, IL. → pages 122

Kiparsky, P. (1968). Linguistic universals and linguistic change. In Bach, E. and Harms, R., editors, *Universals in linguistic theory*, pages 170–202. Holt, Rinehart and Winston, New York. → pages 4

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. In *Proceedings of the National Academy of Sciences*, volume 105, pages 10681–10686. → pages 3

Kobele, G. M. (2006). *Generating copies: an investigation into structural identity in language and grammar*. PhD thesis, University of California, Los Angeles. → pages 15

Koo, H. and Callahan, L. (2012). Tier-adjacency is not a necessary condition for learning phonotactic dependencies. *Language and Cognitive Processes*, 27:1425–1432. → pages 120

Koo, H. and Cole, J. (2006). On learnability and naturalness as constraints on phonological grammar. In Botinis, A., editor, *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics*, pages 165–168, Athens. University of Athens. → pages 11

Koo, H. and Oh, Y.-I. (2013). Beyond tier-based bigrams: an artificial grammar learning study. *Language Sciences*, 38:53–58. → pages 9, 114, 120, 121, 123, 146, 157

Lai, R. (2012). *Domain specificity in learning phonology*. Doctoral dissertation, University of Delaware. → pages 2, 15, 80, 84, 92

Levelt, C. C. (2011). Consonant harmony in child language. In van Oostendorp, M., Ewen, C. J., Hume, E., and Rice, K., editors, *Blackwell companion to phonology*, volume 3, pages 1691–1716. Blackwell, Oxford. → pages 8

Lombardi, L. (1999). Positional faithfulness and voicing assimilation in optimality theory. *Natural Language and Linguistic Theory*, 17:267–302. → pages 16, 153

McCarthy, J. J. (2003). OT constraints are categorical. *Phonology*, 20:75–138. → pages 13

McCarthy, J. J. (2007). Consonant harmony via correspondence: evidence from Chumash. In Bateman, L., O'Keefe, M., Reilly, E., and Werle, A., editors, *Papers in Optimality Theory III (University of Massachusetts Occasional Papers in Linguistics 32)*, pages 223–238. GLSA Publications, University of Massachusetts, Amherst, MA. → pages 140

McMullin, K. (2013). Learning consonant harmony in artificial languages. In *Proceedings of the 2013 Meeting of the Canadian Linguistic Association*. → pages 73

McMullin, K. and Allen, B. H. (2015). Phonotactic learning and the conjunction of Tier-based Strictly Local languages. Paper presented at the 89th Annual Meeting of the Linguistic Society of America. Portland, OR. → pages 137

McMullin, K. and Hansson, G. Ó. (2014). Locality in long-distance phonotactics: evidence for modular learning. In Iyer, J. and Kusmer, L., editors, *Proceedings of the 44th meeting of the North Eastern Linguistic Society*, volume 2, pages 1–14, Amherst. GLSA Publications, University of Massachusetts. → pages 11, 17, 24, 25, 39, 46, 73, 80, 84, 149, 154

McNaughton, R. and Papert, S. A. (1971). *Counter-free automata*. MIT Press, Cambridge, MA. → pages 15, 80, 154

Moreton, E. (2008). Analytic bias and phonological typology. *Phonology*, 25:83–127. → pages 2, 3, 5, 157

Moreton, E. (2012). Inter- and intra-dimensional dependencies in implicit phonotactic learning. *Journal of Memory and Language*, 67:165–183. → pages 2, 3, 5, 157

Moreton, E. and Pater, J. (2012a). Structure and substance in artificial phonology learning. Part I: Structure. *Language and Linguistics Compass*, 6(11):686–701. → pages 5, 6

Moreton, E. and Pater, J. (2012b). Structure and substance in artificial phonology learning. Part II: Substance. *Language and Linguistics Compass*, 6(11):702–718. → pages 5, 6

Morley, R. (2015). Can phonological universals be emergent? modeling the space of sound change, lexical distribution, and hypothesis selection. *Language*, 91(2):e40–e70. → pages 2

Nevins, A. (2010). *Locality in vowel harmony*. MIT Press, Cambridge, MA. → pages 6, 158

Newport, E. L. and Aslin, R. N. (2004). Learning at a distance I: statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48:127–162. → pages 1, 9

Ní Chiosáin, M. and Padgett, J. (2001). Markedness, segment realization, and locality in spreading. In Lombardi, L., editor, *Segmental phonology in optimality theory*, pages 118–156. Cambridge University Press, Cambridge. → pages 6

Odden, D. (1994). Adjacency parameters in phonology. *Language*, 70:289–330. → pages 6, 7, 9, 10, 12, 17, 21, 22, 23, 25, 59, 85, 113

Ohala, J. J. (1993). The phonetics of sound change. In Jones, C., editor, *Historical linguistics: Problems and perspectives*, pages 237–278. Longman, London. → pages 5

Pater, J. and Tessier, A.-M. (2003). Phonotactic knowledge and the acquisition of alternations. In Solé, M., Recasens, D., and Romero, J., editors, *Proceedings of the 15th International Congress on Phonetic Sciences*, pages 1177–1180, Barcelona. → pages 45

Payne, A. (2014). Dissimilation as a subsequential process. In Iyer, J. and Kusmer, L., editors, *Proceedings of the 44th meeting of the North Eastern Linguistic Society*, volume 1, pages 79–90, Amherst. GLSA Publications, University of Massachusetts. → pages 15, 79, 158

Poser, W. J. (1982). Phonological representations and action-at-a-distance. In van der Hulst, H. and Smith, N., editors, *The structure of phonological representations*, volume 2, pages 121–158. Foris, Dordrecht. → pages 6, 140

Poser, W. J. (1993). Are strict cycle effects derivable? In Hargus, S. and Kaisse, E. M., editors, *Studies in lexical phonology*, pages 315–321. Academic Press, New York. → pages 140

Prince, A. and Smolensky, P. (2004). *Optimality Theory: constraint interaction in generative grammar*. Blackwell, Malden, MA. [Originally distributed 1993 as Technical Report RuCCS-TR-2/CU-CS-696-93, Rutgers Center for Cognitive Science, Rutgers University, and Department of Cognitive Science, University of Colorado at Boulder; revised 2002 version as ROA-537]. → pages 13, 139, 151

Prince, A. and Tesar, B. (2004). Learning phonotactic distributions. In Kager, R., Pater, J., and Zonneveld, W., editors, *Fixing priorities: constraints in phonological acquisition*, pages 245–291. Cambridge University Press, Cambridge. → pages 44

Pulleyblank, D. (2002). Harmony drivers: no disagreement allowed. In Larson, J. and Paster, M., editors, *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society*, pages 249–267. Berkeley Linguistics Society, Berkeley, CA. → pages 6, 9, 86, 142

Pycha, A., Nowak, P., Shin, E., and Shosted, R. (2003). Phonological rule-learing and its implications for a theory of vowel harmony. In Tsujimura, M. and Garding, G., editors, *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, pages 423–435, Somerville, MA. Cascadilla Press. → pages 5

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. → pages 32

Rabin, M. O. and Scott, D. (1959). Finite automata and their decision problems. *IBM Journal of Research and Development*, 3:114–125. → pages 14, 79

Rafferty, A. N., Griffiths, T. L., and Ettlinger, M. (2013). Greater learnability is not sufficient to produce cultural universals. *Cognition*, 129:70–87. → pages 3

Reber, R. and Perruchet, P. (2003). The use of control groups in artificial grammar learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 56A:91–115. → pages 6

Riggle, J. (2004). *Generation, recognition and learning in finite-state optimality theory*. Doctoral dissertation, University of California, Los Angeles. → pages 13

Rogers, J., Heinz, J., Bailey, G., Edlefsen, M., Visscher, M., Wellcome, D., and Wibel, S. (2010). On languages piecewise testable in the strict sense. In Ebert, C., Jäger, G., and Michaelis, J., editors, *The mathematics of language*, pages 255–265. Springer, Berlin. → pages 15, 80, 81, 154

Rogers, J. and Pullum, G. K. (2011). Aural pattern recognition experiments and the subregular hierarchy. *Journal of Logic, Language and Information*, 20(3):329–342. → pages 15, 80, 91, 92, 154

Rose, S. and Walker, R. (2004). A typology of consonant agreement as correspondence. *Language*, 80:475–531. → pages ii, 6, 8, 10, 11, 14, 18, 20, 21, 22, 39, 41, 48, 51, 52, 53, 54, 78, 79, 84, 151

Scott-Phillips, T. and Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences*, 14:411–417. → pages 3

Shaw, P. A. (1991). Consonant harmony systems: the special status of coronal harmony. In Paradis, C. and Prunet, J.-F., editors, *The special status of coronals: internal and external evidence*, pages 125–158. Academic Press, San Diego, CA. → pages 17, 113

Shieber, S. (1985). Evidence against the context-freeness of natural language. *Linguistics and Philosophy*, 8:333–343. → pages 15

Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry*, 21:720–731. → pages 44

Steriade, D. (1987a). Locality conditions and feature geometry. In McDonough, J. and Plunkett, B., editors, *Proceedings of the 17th meeting of the North Eastern Linguistic Society*, pages 595–617. GLSA Publications, University of Massachusetts, Amherst, MA. → pages 6, 12, 118, 119

Steriade, D. (1987b). Redundant values. In Bosch, A., Need, B., and Schiller, E., editors, *CLS 23: Papers from the parasession on autosegmental and metrical phonology*, pages 339–362. Chicago Linguistic Society, Chicago, IL. → pages 6

Suzuki, K. (1998). *A typological investigation of dissimilation*. Doctoral dissertation, University of Arizona. → pages 8, 142

Tesar, B. and Smolensky, P. (2000). *Learnability in Optimality Theory*. MIT Press, Cambridge, MA. → pages 13, 44

van de Weijer, J. (2014). The origin of OT constraints. *Lingua*, 142:66–75. → pages 14

van der Hulst, H. and van de Weijer, J. (1995). Vowel harmony. In Goldsmith, J., editor, *The handbook of phonological theory*, pages 495–531. Blackwell, Oxford. → pages 158

Walker, R. (2000a). Long-distance consonantal identity effects. In Billerey, R. and Lillehaugen, B. D., editors, *Proceedings of the 19th West Coast Conference on Formal Linguistics*, pages 532–545. Cascadilla Press, Somerville, MA. → pages 14, 39, 78, 79, 151

Walker, R. (2000b). *Nasalization, neutral segments, and opacity effects*. Garland, New York. → pages 158

Walker, R. (2000c). Yaka nasal harmony: spreading or segmental correspondence? In Conathan, L., Good, J., Kavitskaya, D., Wulf, A., and Yu, A. C. L., editors, *Proceedings of the 26th Annual Meeting of the Berkeley Linguistics Society*, pages 321–332. Berkeley Linguistics Society, Berkeley, CA. → pages 14, 39, 78, 79, 151

Walker, R. (2001). Consonantal correspondence. In Kirchner, R., Pater, J., and Wikeley, W., editors, *Proceedings of the Workshop on the Lexicon in Phonetics and Phonology (Papers in Experimental and Theoretical Linguistics 6)*, pages 73–84, Edmonton. Department of Linguistics, University of Alberta. → pages 39

Walker, R., Byrd, D., and Mpiranya, F. (2008). An articulatory view of Kinyarwanda coronal harmony. *Phonology*, 25:499–535. → pages 12, 88, 114, 116, 117, 122

Walker, R. and Mpiranya, F. (2005). On triggers and opacity in coronal harmony. In *Proceedings of the Berkeley Linguistic Society*, volume 31, pages 383–394. → pages 12, 88, 114, 117, 122

Warker, J. A., Xu, Y., Dell, G. S., and Fisher, C. (2009). Speech errors reflect the phonotactic constraints in recently spoken syllables, but not in recently heard syllables. *Cognition*, 112:81–96. → pages 28

Watkins, C. (1970). A case of non-chronological rule insertion. *Linguistic Inquiry*, 1:525–527. → pages 12, 118, 119

Wilson, C. (2003). Experimental investigation of phonological naturalness. In Tsujimura, M. and Garding, G., editors, *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, pages 533–546, Somerville, MA. Cascadilla Press. → pages 5, 157

Wilson, C. (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science*, 30:945–982. → pages 3, 5, 93

Appendix A

Full List of Stimuli Used in Experiments

Table A.1: List of stimuli used in the practice phase.

Stem	Past	Future
bigo	bigoli	bigoru
deto	detoli	detoru
gone	goneli	goneru
kumu	kumuli	kumuru
mebi	mebili	mebiru
nipu	nipuli	nipuru
pudi	pudili	pudiru
toke	tokeli	tokeru

Table A.2: Training stimuli with liquids at "Medium-range"

M-Stem	M-Harm-Past	M-Harm-Fut	M-Diss-Past	M-Diss-Fut
polipu	polipuli	poripuru	poripuli	polipuru
pilede	piledeli	pirederu	piredeli	pilederu
bilono	bilonoli	bironoru	bironoli	bilonoru
neluki	nelukili	nerukiru	nerukili	nelukiru
belibu	belibuli	beriburu	beribuli	beliburu
pelege	pelegeli	peregeru	peregeli	pelegeru
tilipe	tilipeli	tiriperu	tiripeli	tiliperu
noleni	nolenili	noreniru	norenili	noleniru
nuloto	nulotoli	nurotoru	nurotoli	nulotoru
giluko	gilukoli	girukoru	girukoli	gilukoru
melomi	melomili	meromiru	meromili	melomiru
kolugu	koluguli	koruguru	koruguli	koluguru
berigi	beligili	berigiru	berigili	beligiru
mureke	mulekeli	murekeru	murekeli	mulekeru
moropo	molopoli	moroporu	moropoli	moloporu
gurubo	guluboli	guruboru	guruboli	guluboru
nuronu	nulonuli	nuronuru	nuronuli	nulonuru
girudi	giludili	girudiru	girudili	giludiru
guritu	gulituli	gurituru	gurituli	gulituru
piredu	pileduli	pireduru	pireduli	pileduru
birobe	bilobeli	biroberu	birobeli	biloberu
neruti	nelutili	nerutiru	nerutili	nelutiru
tirime	tilimeli	tirimeru	tirimeli	tilimeru
peremo	pelemoli	peremoru	peremoli	pelemoru
gulidu	guliduli	guriduru	guriduli	guliduru
mulegu	muleguli	mureguru	mureguli	muleguru
molobi	molobili	morobiru	morobili	molobiru
gulune	guluneli	guruneru	guruneli	guluneru
bolipi	bolipili	boripiru	boripili	bolipiru

Continued on next page

Table A.2 – Medium-range training (*Continued from previous page*)

M-Stem	M-Harm-Past	M-Harm-Fut	M-Diss-Past	M-Diss-Fut
keluko	kelukoli	kerukoru	kerukoli	kelukoru
delito	delitoli	deritoru	deritoli	delitoru
dilemo	dilemoli	diremoru	diremoli	dilemoru
nilobu	nilobuli	niroburu	nirobuli	niloburu
tilute	tiluteli	tiruteru	tiruteli	tiluteru
telede	teledeli	terederu	teredeli	telederu
bulomi	bulomili	buromiru	buromili	bulomiru
poriku	polikuli	porikuru	porikuli	polikuru
norego	nolegoli	noregoru	noregoli	nolegoru
meroni	melonili	meroniru	meronili	meloniru
korupe	kolupeli	koruperu	korupeli	koluperu
turebe	tulebeli	tureberu	turebeli	tuleberu
gorodo	golodoli	gorodoru	gorodoli	golodoru
deriki	delikili	derikiru	derikili	delikiru
direno	dilenoli	direnoru	direnoli	dilenoru
buromu	bulomuli	buromuru	buromuli	bulomuru
muruge	mulugeli	murugeru	murugeli	mulugeru
kuripu	kulipuli	kuripuru	kuripuli	kulipuru
tiruti	tilutili	tirutiru	tirutili	tilutiru
gilipo	gilipoli	giriporu	giripoli	giliporu
koledi	koledili	korediru	koredili	kolediru
nelogi	nelogili	nerogiru	nerogili	nelogiru
molunu	molunuli	morunuru	morunuli	molunuru
golobo	goloboli	goroboru	goroboli	goloboru
muluke	mulukeli	murukeru	murukeli	mulukeru
kulipo	kulipoli	kuriporu	kuripoli	kuliporu
tulegi	tulegili	turegiru	turegili	tulegiru
bilotu	bilotuli	biroturu	birotuli	biloturu
pelume	pelumeli	perumeru	perumeli	pelumeru
tolitu	tolituli	torituru	torituli	tolituru

Continued on next page

Table A.2 – Medium-range training (*Continued from previous page*)

M-Stem	M-Harm-Past	M-Harm-Fut	M-Diss-Past	M-Diss-Fut
mileme	milemeli	miremeru	miremeli	milemeru
girige	giligeli	girigeru	girigeli	giligeru
tereno	telenoli	terenoru	terenoli	telenoru
niropu	nilopuli	niropuru	niropuli	nilopuru
kerude	keludeli	keruderu	kerudeli	keluderu
boriki	bolikili	borikiru	borikili	bolikiru
korebu	kolebuli	koreburu	korebuli	koleburu
geridi	gelidili	geridiru	geridili	gelidiru
pureto	puletoli	puretoru	puretoli	puletoru
nerobo	neloboli	neroboru	neroboli	neloboru
morumi	molumili	morumiru	morumili	molumiru
toronu	tolonuli	toronuru	toronuli	tolonuru
duruke	dulukeli	durukeru	durukeli	dulukeru
gelibe	gelibeli	geriberu	geribeli	geliberu
kelego	kelegoli	keregoru	keregoli	kelegoru
buloku	bulokuli	burokuru	burokuli	bulokuru
diluni	dilunili	diruniru	dirunili	diluniru
dolepe	dolepeli	doreperu	dorepeli	doleperu
melodo	melodoli	merodoru	merodoli	melodoru
kilibi	kilibili	kiribiru	kiribili	kilibiru
puleti	puletili	puretiru	puretili	puletiru
tolodo	tolodoli	torodoru	torodoli	tolodoru
poluku	polukuli	porukuru	porukuli	polukuru
dulimu	dulimuli	durimuru	durimuli	dulimuru
dulune	duluneli	duruneru	duruneli	duluneru
duribi	dulibili	duribiru	duribili	dulibiru
mireko	milekoli	mirekoru	mirekoli	milekoru
burogu	bulogulі	buroguru	buroguli	buloguru
dirudu	diluduli	diruduru	diruduli	diluduru
kirine	kilineli	kirineru	kirineli	kilineru

Continued on next page

Table A.2 – Medium-range training (*Continued from previous page*)

M-Stem	M-Harm-Past	M-Harm-Fut	M-Diss-Past	M-Diss-Fut
perupi	pelupili	perupiru	perupili	pelupiru
torite	toliteli	toriteru	toriteli	toliteru
kerepi	kelepili	kerepiru	kerepili	kelepiru
birogo	bilogoli	birogoru	birogoli	bilogoru
porumo	polumoli	porumoru	porumoli	polumoru
dorete	doleteli	doreteru	doreteli	doleteru
meromu	melomuli	meromuru	meromuli	melomuru

Table A.3: Training stimuli with liquids at "Short-range"

S-Stem	S-Harm-Past	S-Harm-Fut	S-Diss-Past	S-Diss-Fut
pupoli	pupolili	puporiru	puporili	pupoliru
depile	depileli	depireru	depireli	depileru
nobilo	nobiloli	nobiroru	nobiroli	nobiloru
kinelu	kineluli	kineruru	kineruli	kineluru
bubeli	bubelili	buberiru	buberili	bubeliru
gepele	gepeleli	gepereru	gepereli	gepeleru
petili	petilili	petiriru	petirili	petiliru
ninole	ninoleli	ninoreru	ninoreli	ninoleru
tonulo	tonuloli	tonuroru	tonuroli	tonuloru
kogilu	kogiluli	kogiruru	kogiruli	kogiluru
mimelo	mimeloli	mimeroru	mimeroli	mimeloru
gukolu	gukoluli	gukoruru	gukoruli	gukoluru
giberi	gibelili	giberiru	giberili	gibeliru
kemure	kemuleli	kemureru	kemureli	kemuleru
pomoro	pomololi	pomororu	pomoroli	pomoloru
boguru	bogululi	boguruu	boguruli	boguluru
nunuro	nunuloli	nunuroru	nunuroli	nunuloru
digiru	digiluli	digiruru	digiruli	digiluru
tuguri	tugulili	tuguriru	tugurili	tuguliru
dupire	dupileli	dupireru	dupireli	dupileru
bebiro	bebiloli	bebiroru	bebiroli	bebiloru
tineru	tineluli	tineruru	tineruli	tineluru
metiri	metilili	metiriru	metirili	metiliru
mopere	mopeleli	mopereru	mopereli	mopeleru
duguli	dugulili	duguriru	dugurili	duguliru
gumule	gumuleli	gumureru	gumureli	gumuleru
bimolo	bimololi	bimororu	bimoroli	bimoloru
negulu	neguluii	neguruu	neguruli	neguluru
piboli	pibolili	piboriru	piborili	piboliru

Continued on next page

Table A.3 – Short-range training (*Continued from previous page*)

S-Stem	S-Harm-Past	S-Harm-Fut	S-Diss-Past	S-Diss-Fut
kokelu	kokeluli	kokeruru	kokeruli	kokeluru
todeli	todelili	toderiru	toderili	todeliru
modile	modileli	modireru	modireli	modileru
bunilo	buniloli	buniroru	buniroli	buniloru
tetilu	tetiluli	tetiruru	tetiruli	tetiluru
detele	deteleli	detereru	detereli	deteleru
mibulo	mibuloli	miburoru	miburoli	mibuloru
kupori	kupolili	kuporiru	kuporili	kupoliru
gonore	gonoleli	gonoreru	gonoreli	gonoleru
nimero	nimeloli	nimeroru	nimeroli	nimeloru
pekoru	pekoluli	pekoruru	pekoruli	pekoluru
beture	betuleli	betureru	betureli	betuleru
dogoro	dogololi	dogororu	dogoroli	dogoloru
kideri	kidelili	kideriru	kiderili	kideliru
nodire	nodileli	nodireru	nodireli	nodileru
muburo	mubuloli	muburoru	muburoli	mubuloru
gemuru	gemululi	gemururu	gemuruli	gemuluru
pukuri	pukulili	pukuriru	pukurili	pukuliru
titiru	titiluli	titiruru	titiruli	titiluru
pogili	pogilili	pogiriru	pogirili	pogiliru
dikole	dikoleli	dikoreru	dikoreli	dikoleru
ginelo	gineloli	gineroru	gineroli	gineloru
numolu	numoluli	numoruru	numoruli	numoluru
bogolo	bogololi	bogororu	bogoroli	bogoloru
kemulu	kemululi	kemururu	kemuruli	kemuluru
pokuli	pokulili	pokuriru	pokurili	pokuliru
gitule	gituleli	gitureru	gitureli	gituleru
tubilo	tubiloli	tubiroru	tubiroli	tubiloru
mepelu	mepeluli	meperuru	meperuli	mepeluru
tutoli	tutolili	tutoriru	tutorili	tutoliru

Continued on next page

Table A.3 – Short-range training (*Continued from previous page*)

S-Stem	S-Harm-Past	S-Harm-Fut	S-Diss-Past	S-Diss-Fut
memile	memileli	memireru	memireli	memileru
gegiri	gegilili	gegiriru	gegirili	gegiliru
notere	noteleli	notereru	notereli	noteleru
puniro	puniloli	puniroru	puniroli	puniloru
dekeru	dekeluli	dekeruru	dekeruli	dekeluru
kibori	kibolili	kiboriru	kiborili	kiboliru
bukore	bukoleli	bukoreru	bukoreli	bukoleru
digeri	digelili	digeriru	digerili	digeliru
topure	topuleli	topureru	topureli	topuleru
bonero	boneloli	boneroru	boneroli	boneloru
mimoru	mimoluli	mimoruru	mimoruli	mimoluru
nutoro	nutololi	nutororu	nutoroli	nutoloru
keduru	kedululi	keduruu	keduruli	keduluru
begeli	begelili	begeriru	begerili	begeliru
gokele	gokeleli	gokereru	gokereli	gokeleru
kubulo	kubuloli	kuburoru	kuburoli	kubuloru
nidilu	nidiluli	nidiruru	nidiruli	nidiluru
pedole	pedoleli	pedoreru	pedoreli	pedoleru
domelo	domeloli	domeroru	domeroli	domeloru
bikili	bikilili	bikiriru	bikirili	bikiliru
tipule	tipuleli	tipureru	tipureli	tipuleru
dotolo	dotololi	dotororu	dotoroli	dotoloru
kupolu	kupoluli	kuporuru	kuporuli	kupoluru
muduli	mudulili	muduriru	mudurili	muduliru
nedulu	nedululi	nedururu	neduruli	neduluru
biduri	bidulili	biduriru	bidurili	biduliru
komire	komileli	komireru	komireli	komileru
guburo	gubuloli	guburoru	guburoli	gubuloru
dudiru	dudiluli	dudiruru	dudiruli	dudiluru
nekiri	nekilili	nekiriru	nekirili	nekiliru

Continued on next page

Table A.3 – Short-range training (*Continued from previous page*)

S-Stem	S-Harm-Past	S-Harm-Fut	S-Diss-Past	S-Diss-Fut
piperu	pipeluli	piperuru	piperuli	pipeluru
tetori	tetolili	tetoriru	tetorili	tetoliru
pikere	pikeleli	pikereru	pikereli	pikeleru
gobiro	gobiloli	gobiroru	gobiroli	gobiloru
moporu	mopoluli	moporuru	moporuli	mopoluru
tedore	tedoleli	tedoreru	tedoreli	tedoleru
mumero	mumeloli	mumeroru	mumeroli	mumeloru

Table A.4: Training stimuli with no liquids.

Stem	Past	Future
tikemu	tikemuli	tikemuru
kibupi	kibupili	kibupiru
pupugu	pupuguli	pupuguru
gonuni	gonunili	gonuniru
bipobe	bipobeli	bipoberu
tepobi	tepobili	tepobiru
tomeku	tomekuli	tomekuru
pibogo	pibogoli	pibogoru
nekine	nekineli	nekineru
mutumu	mutumuli	mutumuru
dubope	dubopeli	duboperu
degiti	degitili	degitiru
kukedo	kukedoli	kukedoru
nomene	nomeneli	nomeneru
gegebi	gegebili	gegebiru
butopi	butopili	butopiru
dodigo	dodigoli	dodigoru
nimimo	nimimoli	nimimoru
pededu	pededuli	pededuru
minoko	minokoli	minokoru
gutudo	gutudoli	gutudoru
mogiku	mogikuli	mogikuru
konute	konuteli	konuteru
bedite	bediteli	bediteru
podoge	podogeli	podogeru
gibipe	gibipeli	gibiperu
topidu	topiduli	topiduru
potetu	potetuli	poteturu
bumumo	bumumoli	bumumoru

Continued on next page

Table A.4 – CVCVCV-<u>L</u>V training
(*Continued from previous page*)

Stem	Past	Future
dimumi	dimumili	dimumiru
botini	botinili	botiniru
gipebu	gipebuli	gipeburu
denenu	denenuli	denenuru
nupidi	nupidili	nupidiru
bokuno	bokunoli	bokunoru
nonegu	noneguli	noneguru
kemoti	kemotili	kemotiru
digupo	digupoli	diguporu
pubigi	pubigili	pubigiru
medoto	medotoli	medotoru
tuniki	tunikili	tunikiru
kekoke	kekokeli	kekokeru
megobe	megobeli	megoberu
gebepu	gebepuli	gebepuru
migode	migodeli	migoderu
nukuko	nukukoli	nukukoru
kidubo	kiduboli	kiduboru
tuteme	tutemeli	tutemeru
pemoti	pemotili	pemotiru
tetobu	tetobuli	tetoburu
tetige	tetigeli	tetigeru
begiku	begikuli	begikuru
pipeto	pipetoli	pipetoru
gokine	gokineli	gokineru
monube	monubeli	monuberu
dobemu	dobemuli	dobemuru
nituki	nitukili	nitukiru
mupudu	mupuduli	mupuduru

Continued on next page

Table A.4 – CVCVCV-LV training
(*Continued from previous page*)

Stem	Past	Future
pibupi	pibupili	pibupiru
doduno	dodunoli	dodunoru
kugumi	kugumili	kugumiru
kugeko	kugekoli	kugekoru
nemide	nemideli	nemideru
mubope	mubopeli	muboperu
bikote	bikoteli	bikoteru
tidido	tididoli	tididoru
gomepu	gomepuli	gomepuru
duponi	duponili	duponiru
kodegu	kodeguli	kodeguru
nenobi	nenobili	nenobiru
gukego	gukegoli	gukegoru
binimo	binimoli	binimoru
mipede	mipedeli	mipederu
pegono	pegonoli	pegonoru
kikuge	kikugeli	kikugeru
dibigi	dibigili	dibigiru
bukoke	bukokeli	bukokeru
ninipu	ninipuli	ninipuru
bemubo	bemuboli	bemuboru
kedetu	kedetuli	kedeturu
bobeki	bobekili	bobekiru
gepunu	gepunuli	gepunuru
podupo	podupoli	poduporu
metopo	metopoli	metoporu
nodobu	nodobuli	nodoburu
titeme	titemeli	titemeru
kotike	kotikeli	kotikeru

Continued on next page

Table A.4 – CVCVCV-<u>L</u>V training
(*Continued from previous page*)

Stem	Past	Future
nubutu	nubutuli	nubuturu
momime	momimeli	momimeru
ginoto	ginotoli	ginotoru
tunenu	tunenuli	tunenuru
gukidi	gukidili	gukidiru
pumogi	pumogili	pumogiru
togudi	togudili	togudiru
degemi	degemili	degemiru
dupibo	dupiboli	dupiboru

Table A.5: List of stimuli used in testing phase.

Distance	Stem	“l” option	“r” option
Short	dotile	dotileli	dotireli
Short	tipoli	tipolili	tiporili
Short	bibolo	bibololi	biboroli
Short	pudele	pudeleli	pudereli
Short	guneli	gunelili	gunerili
Short	momilu	momiluli	momiruli
Short	negulu	negululi	neguruli
Short	kekulo	kekuloli	kekuroli
Short	pidole	pidoleru	pidoreru
Short	nonolu	nonoluru	nonoruru
Short	tepilo	tepiloru	tepiroru
Short	gigili	gigiliru	gigiriru
Short	mukelu	mukeluru	mukeruru
Short	detule	detuleru	detureru
Short	komuli	komuliru	komuriru
Short	bubelo	bubeloru	buberoru
Short	gegori	gegolili	gegorili
Short	kidure	kiduleli	kidureli
Short	dutere	duteleli	dutereli
Short	popero	popeloli	poperoli
Short	nibiru	nibiluli	nibiruli
Short	memoru	memoluli	memoruli
Short	bonuro	bonuloli	bonuroli
Short	tukiri	tukilili	tukirili
Short	mipuru	mipuluru	mipururu
Short	ditore	ditoleru	ditoreru
Short	pemeri	pemeliru	pemeriru
Short	goniro	goniloru	goniroru
Short	kudire	kudileru	kudireru

Continued on next page

Table A.5 – Testing stimuli
(*Continued from previous page*)

Distance	Stem	“l” option	“r” option
Short	nuburi	nubuliru	nuburiru
Short	tokoro	tokoloru	tokororu
Short	begeru	begeluru	begeruru
Medium	beliki	belikili	berikili
Medium	dilopo	dilopoli	diropoli
Medium	molutu	molutuli	morutuli
Medium	pelemi	pelemili	peremili
Medium	kilono	kilonoli	kironoli
Medium	gulibe	gulibeli	guribeli
Medium	tuluge	tulugeli	turugeli
Medium	noledu	noleduli	noreduli
Medium	pilepe	pileperu	pireperu
Medium	muluto	mulutoru	murutoru
Medium	nulimu	nulimuru	nurimuru
Medium	tolone	toloneru	toroneru
Medium	golobi	golobiru	gorobiru
Medium	keleku	kelekuru	kerekuru
Medium	deludo	deludoru	derudoru
Medium	bilegi	bilegiru	biregiru
Medium	burike	bulikeli	burikeli
Medium	dorupi	dolupili	dorupili
Medium	merenu	melenuli	merenuli
Medium	puremo	pulemoli	puremoli
Medium	korogo	kologoli	korogoli
Medium	nirobu	nilobuli	nirobuli
Medium	tiriti	tilitili	tiritili
Medium	gerude	geludeli	gerudeli
Medium	porodi	polodiru	porodiru
Medium	mirete	mileteru	mireteru

Continued on next page

Table A.5 – Testing stimuli

(*Continued from previous page*)

Distance	Stem	“l” option	“r” option
Medium	girupu	gilupuru	girupuru
Medium	teriko	telikoru	terikoru
Medium	nereme	nelemeru	neremeru
Medium	kuroni	kuloniru	kuroniru
Medium	duribo	duliboru	duriboru
Medium	borugu	boluguru	boruguru
Long	letubi	letubili	retubili
Long	linode	linodeli	rinodeli
Long	limegu	limeguli	rimeguli
Long	lugupi	lugupili	rugupili
Long	ledimo	ledimoli	redimoli
Long	lokenu	lokenuli	rokenuli
Long	lipoke	lipokeli	ripokeli
Long	lebito	lebitoli	rebitoli
Long	lunedo	lunedoru	runedoru
Long	lotiku	lotikuru	rotikuru
Long	lokite	lokiteru	rokiteru
Long	lemogo	lemogoru	remogoru
Long	lugoni	lugoniru	rugoniru
Long	lipube	lipuberu	ripuberu
Long	lobupu	lobupuru	robupuru
Long	ludemi	ludemiru	rudemiru
Long	rupimu	lupimuli	rupimuli
Long	ronupe	lonupeli	ronupeli
Long	romuge	lomugeli	romugeli
Long	rebeti	lebetili	rebetili
Long	ruteki	lutekili	rutekili
Long	rikono	likonoli	rikonoli
Long	rodobo	lodoboli	rodoboli

Continued on next page

Table A.5 – Testing stimuli
(*Continued from previous page*)

Distance	Stem	“l” option	“r” option
Long	rugidu	lugiduli	rugiduli
Long	regedi	legediru	regediru
Long	ritoko	litokoru	ritokoru
Long	ribopo	liboporu	riboporu
Long	rumibu	lumiburu	rumiburu
Long	renitu	lenituru	renituru
Long	rodume	lodumeru	rodumeru
Long	rekune	lekuneru	rekuneru
Long	ripegi	lipegiru	ripegiru

Appendix B

Statistical Analyses

B.1 Experiment 1 (16 Subjects)

Table B.1: Summary of the fixed effects portion of the mixed-effects logistic regression for Experiment 1 with Short-range baseline (N = 4518; log-likelihood = –2235.4)

Coefficient	Estimate	*SE*	*Pr*(>|*z*|)
Intercept	−1.13651	0.26127	< 0.0001
Harmony Faithful	2.47228	0.28800	< 0.0001
Harmony Second	−0.50501	0.12784	< 0.0001
Medium-range	0.09456	0.16467	0.5658
Long-range	−0.18878	0.16318	0.2473
S-Harm	1.28500	0.29777	< 0.0001
M-Harm	1.10671	0.30310	0.0003
Medium-range × S-Harm	−1.02310	0.22904	< 0.0001
Long-range × S-Harm	−0.89644	0.22821	< 0.0001
Medium-range × M-Harm	0.22645	0.23169	0.3284
Long-range × M-Harm	−0.29497	0.22715	0.1941

Table B.2: Summary of the fixed effects portion of the mixed-effects logistic regression for Experiment 1 with Medium-range baseline (N = 4518; log-likelihood = –2235.4)

Coefficient	Estimate	SE	$Pr(>\|z\|)$
Intercept	-1.04300	0.26124	< 0.0001
Harmony Faithful	2.47209	0.28791	< 0.0001
Harmony Second	-0.50513	0.12783	< 0.0001
Short-range	-0.09291	0.16463	0.5725
Long-range	-0.28145	0.16398	0.0861
S-Harm	0.26217	0.29489	0.3740
M-Harm	1.33476	0.30544	< 0.0001
Short-range × S-Harm	1.02171	0.22899	< 0.0001
Long-range × S-Harm	0.12460	0.22549	0.5805
Short-range × M-Harm	-0.22857	0.23162	0.3237
Long-range × M-Harm	-0.52396	0.23064	0.0231

Table B.3: Summary of the fixed effects portion of the mixed-effects logistic regression for Experiment 1 with Long-range baseline (N = 4518; log-likelihood = –2235.4)

Coefficient	Estimate	SE	$Pr(>\|z\|)$
Intercept	-1.3242	0.2614	< 0.0001
Harmony Faithful	2.4724	0.2879	< 0.0001
Harmony Second	-0.5050	0.1278	< 0.0001
Short-range	0.1877	0.1632	0.2500
Medium-range	0.2826	0.1640	0.0848
S-Harm	0.3877	0.2940	0.1872
M-Harm	0.8975	0.3005	0.0071
Short-range × S-Harm	0.8975	0.2282	< 0.0001
Medium-range × S-Harm	-0.1260	0.2255	0.5764
Short-range × M-Harm	0.2964	0.2271	0.1919
Medium-range × M-Harm	0.5225	0.2307	0.0235

B.2 Experiment 1 (12 “Successful Learners”)

Table B.4: Summary of the fixed effects portion of the mixed-effects logistic regression for learners in Experiment 1 with Short-range baseline (N = 3400; log-likelihood = –1491.9)

Coefficient	Estimate	*SE*	*Pr*(>|*z*|)
Intercept	−1.14488	0.23378	< 0.0001
Harmony Faithful	2.43613	0.32511	< 0.0001
Harmony Second	−0.37411	0.12162	0.0021
Medium-range	0.04516	0.18374	0.8058
Long-range	−0.18983	0.18263	0.2986
S-Harm	2.67994	0.26191	< 0.0001
M-Harm	2.48025	0.26095	< 0.0001
Medium-range × S-Harm	−2.10084	0.28427	< 0.0001
Long-range × S-Harm	−2.18634	0.28641	< 0.0001
Medium-range × M-Harm	0.07936	0.30243	0.7930
Long-range × M-Harm	−1.18575	0.28199	< 0.0001

Table B.5: Summary of the fixed effects portion of the mixed-effects logistic regression for learners in Experiment 1 with Medium-range baseline (N = 3400; log-likelihood = –1491.9)

Coefficient	Estimate	*SE*	$Pr(>\|z\|)$
Intercept	-1.10047	0.23475	< 0.0001
Harmony Faithful	2.43685	0.32674	< 0.0001
Harmony Second	-0.37468	0.12170	0.0021
Short-range	-0.04425	0.18376	0.8097
Long-range	-0.23266	0.18318	0.2041
S-Harm	0.57830	0.22909	0.0116
M-Harm	2.56041	0.26558	< 0.0001
Short-range × S-Harm	2.10158	0.28435	< 0.0001
Long-range × S-Harm	-0.08781	0.26121	0.7368
Short-range × M-Harm	-0.08072	0.30242	0.7895
Long-range × M-Harm	-1.26740	0.28582	< 0.0001

Table B.6: Summary of the fixed effects portion of the mixed-effects logistic regression for learners in Experiment 1 with Long-range baseline (N = 3400; log-likelihood = –1491.9)

Coefficient	Estimate	*SE*	$Pr(>\|z\|)$
Intercept	-1.33227	0.23452	< 0.0001
Harmony Faithful	2.43742	0.32626	< 0.0001
Harmony Second	-0.37446	0.121704	0.0021
Short-range	0.18682	0.18263	0.3063
Medium-range	0.23086	0.18317	0.2075
S-Harm	0.48959	0.22688	0.0309
M-Harm	1.29107	0.26724	< 0.0001
Short-range × S-Harm	2.19090	0.28648	< 0.0001
Medium-range × S-Harm	0.08921	0.26120	0.7327
Short-range × M-Harm	1.18938	0.28196	< 0.0001
Medium-range × M-Harm	1.26937	0.28581	< 0.0001

B.3 Experiment 2 (16 Subjects)

Table B.7: Summary of the fixed effects portion of the mixed-effects logistic regression for Experiment 2 with Short-range baseline (N = 4534; log-likelihood = –2128.3)

Coefficient	Estimate	*SE*	$Pr(>\|z\|)$
Intercept	-0.72714	0.24416	0.0029
Dissimilation Faithful	2.38800	0.29266	< 0.0001
Dissimilation Second	-0.71309	0.13950	< 0.0001
Medium-range	-0.09619	0.16489	0.5596
Long-range	0.18624	0.16338	0.2543
S-Diss	2.51590	0.30450	< 0.0001
M-Diss	1.47592	0.30610	< 0.0001
Medium-range × S-Diss	-2.03036	0.24977	< 0.0001
Long-range × S-Diss	-2.70521	0.25118	< 0.0001
Medium-range × M-Diss	-0.00647	0.23652	0.9782
Long-range × M-Diss	-0.98679	0.23280	< 0.0001

Table B.8: Summary of the fixed effects portion of the mixed-effects logistic regression for Experiment 2 with Medium-range baseline (N = 4534; log-likelihood = –2128.3)

Coefficient	Estimate	*SE*	*Pr*(>|z|)
Intercept	−0.82611	0.24453	0.0007
Dissimilation Faithful	2.38980	0.29295	< 0.0001
Dissimilation Second	−0.71033	0.13961	< 0.0001
Short-range	0.09523	0.16490	0.5636
Long-range	0.28440	0.16423	0.0833
S-Diss	0.48717	0.28413	0.0864
M-Diss	1.47160	0.30499	< 0.0001
Short-range × S-Diss	2.03270	0.24984	< 0.0001
Long-range × S-Diss	−0.67777	0.22487	< 0.0001
Short-range × M-Diss	0.00435	0.23653	0.9853
Long-range × M-Diss	−0.98355	0.23217	< 0.0001

Table B.9: Summary of the fixed effects portion of the mixed-effects logistic regression for Experiment 2 with Long-range baseline (N = 4534; log-likelihood = –2128.3)

Coefficient	Estimate	*SE*	*Pr*(>|z|)
Intercept	−0.5424	0.2431	0.0257
Dissimilation Faithful	2.3903	0.2933	< 0.0001
Dissimilation Second	−0.7096	0.1397	< 0.0001
Short-range	−0.1892	0.1634	0.2469
Medium-range	−0.2843	0.1642	0.0835
S-Diss	−0.1908	0.2829	0.5000
M-Diss	0.4884	0.2997	0.1032
Short-range × S-Diss	2.7103	0.2512	< 0.0001
Medium-range × S-Diss	0.6781	0.2249	0.0026
Short-range × M-Diss	0.9875	0.2328	< 0.0001
Medium-range × M-Diss	0.9830	0.2322	< 0.0001

B.4 Experiment 2 (12 “Successful Learners”)

Table B.10: Summary of the fixed effects portion of the mixed-effects logistic regression for first 12 learners in Experiment 2 with Short-range baseline (N = 3403; log-likelihood = –1564.2)

Coefficient	Estimate	*SE*	$Pr(>\|z\|)$
Intercept	-0.69269	0.24504	0.0047
Dissimilation Faithful	2.16993	0.33560	< 0.0001
Dissimilation Second	-0.57469	0.12261	< 0.0001
Medium-range	-0.04593	0.18393	0.8028
Long-range	0.18743	0.18290	0.3055
S-Diss	2.14799	0.28624	< 0.0001
M-Diss	2.14863	0.30144	< 0.0001
Medium-range × S-Diss	-1.83421	0.27418	< 0.0001
Long-range × S-Diss	-2.35790	0.27517	< 0.0001
Medium-range × M-Diss	0.49996	0.30027	0.0959
Long-range × M-Diss	-1.46357	0.27468	< 0.0001

Table B.11: Summary of the fixed effects portion of the mixed-effects logistic regression for first 12 learners in Experiment 2 with Medium-range baseline (N = 3403; log-likelihood = –1564.2)

Coefficient	Estimate	*SE*	*Pr*(>|*z*|)
Intercept	−0.73727	0.24535	0.0027
Dissimilation Faithful	2.16952	0.33587	< 0.0001
Dissimilation Second	−0.57574	0.12263	< 0.0001
Short-range	0.04554	0.18395	0.8045
Long-range	0.23465	0.18343	0.2008
S-Diss	0.31389	0.26695	0.2397
M-Diss	2.64772	0.31482	< 0.0001
Short-range × S-Diss	1.83424	0.27421	< 0.0001
Long-range × S-Diss	−0.52551	0.25236	0.0373
Short-range × M-Diss	−0.50210	0.30029	0.0945
Long-range × M-Diss	−1.96560	0.28927	< 0.0001

Table B.12: Summary of the fixed effects portion of the mixed-effects logistic regression for first 12 learners in Experiment 2 with Long-range baseline (N = 3403; log-likelihood = –1564.2)

Coefficient	Estimate	*SE*	*Pr*(>|*z*|)
Intercept	−0.5036	0.2444	0.0393
Dissimilation Faithful	2.1752	0.3369	< 0.0001
Dissimilation Second	−0.5773	0.1225	< 0.0001
Short-range	−0.1887	0.1829	0.3024
Medium-range	−0.2337	0.1834	0.2027
S-Diss	−0.2128	0.2657	0.4233
M-Diss	0.6777	0.2808	0.0158
Short-range × S-Diss	2.3585	0.2752	< 0.0001
Medium-range × S-Diss	0.5254	0.2524	0.0374
Short-range × M-Diss	1.4647	0.2747	< 0.0001
Medium-range × M-Diss	1.9622	0.2892	< 0.0001

B.5 Experiment 3 (16 Subjects)

Table B.13: Summary of the fixed effects portion of the mixed-effects logistic regression for Experiment 3 with Short-range baseline (N = 4490; log-likelihood = –2124.8)

Coefficient	Estimate	*SE*	$Pr(>\|z\|)$
Intercept	−1.28430	0.24824	< 0.0001
Harmony Faithful	2.68550	0.31605	< 0.0001
Harmony Second	−0.43627	0.12893	0.0007
Medium-range	0.09698	0.16511	0.5570
Long-range	−0.18805	0.16336	0.2497
S-Harm-M-Faith	2.00380	0.23778	< 0.0001
M-Harm-S-Faith	0.79664	0.22255	0.0003
Medium-range × S-Harm-M-Faith	−1.55412	0.24271	< 0.0001
Long-range × S-Harm-M-Faith	−1.59964	0.24277	< 0.0001
Medium-range × M-Harm-S-Faith	0.12194	0.23500	0.6038
Long-range × M-Harm-S-Faith	−0.24235	0.23050	0.2931

Table B.14: Summary of the fixed effects portion of the mixed-effects logistic regression for Experiment 3 with Medium-range baseline (N = 4490; log-likelihood = –2124.8)

Coefficient	Estimate	*SE*	$Pr(>\|z\|)$
Intercept	-1.18733	0.24887	< 0.0001
Harmony Faithful	2.68610	0.31641	< 0.0001
Harmony Second	-0.43653	0.12893	0.0007
Short-range	-0.09711	0.16512	0.5565
Long-range	-0.28618	0.16432	0.0816
S-Harm-M-Faith	0.44918	0.22104	0.0421
M-Harm-S-Faith	0.91821	0.22555	< 0.0001
Short-range × S-Harm-M-Faith	1.55429	0.24273	< 0.0001
Long-range × S-Harm-M-Faith	-0.04407	0.22868	0.8472
Short-range × M-Harm-S-Faith	-0.12202	0.23500	0.6036
Long-range × M-Harm-S-Faith	-0.36298	0.23268	0.1188

Table B.15: Summary of the fixed effects portion of the mixed-effects logistic regression for Experiment 3 with Long-range baseline (N = 4490; log-likelihood = –2124.8)

Coefficient	Estimate	*SE*	$Pr(>\|z\|)$
Intercept	-1.4733	0.2481	< 0.0001
Harmony Faithful	2.6844	0.3160	< 0.0001
Harmony Second	-0.4364	0.1290	0.0007
Short-range	0.1896	0.1633	0.2458
Medium-range	0.2874	0.1643	0.0802
S-Harm-M-Faith	0.4062	0.2173	0.0616
M-Harm-S-Faith	0.5556	0.2179	0.0108
Short-range × S-Harm-M-Faith	1.5977	0.2428	< 0.0001
Medium-range × S-Harm-M-Faith	0.0435	0.2287	0.8491
Short-range × M-Harm-S-Faith	0.2411	0.2305	0.2955
Medium-range × M-Harm-S-Faith	0.3625	0.2327	0.1193

B.6 Experiment 4 (16 Subjects)

Table B.16: Summary of the fixed effects portion of the mixed-effects logistic regression for Experiment 4 with Short-range baseline (N = 4518; log-likelihood = –2253.0)

Coefficient	Estimate	*SE*	$Pr(>\|z\|)$
Intercept	−0.83807	0.18363	< 0.0001
Dissimilation Faithful	2.56905	0.28066	< 0.0001
Dissimilation Second	−0.57949	0.11888	< 0.0001
Medium-range	−0.09637	0.16470	0.5585
Long-range	0.18660	0.16301	0.2523
S-Diss-M-Faith	1.93214	0.20439	< 0.0001
M-Diss-S-Faith	0.06160	0.18875	0.7442
Medium-range × S-Diss-M-Faith	−1.63444	0.23878	< 0.0001
Long-range × S-Diss-M-Faith	−2.04029	0.23802	< 0.0001
Medium-range × M-Diss-S-Faith	0.51938	0.22372	0.0203
Long-range × M-Diss-S-Faith	−0.40628	0.22213	0.0674

Table B.17: Summary of the fixed effects portion of the mixed-effects logistic regression for Experiment 4 with Medium-range baseline (N = 4518; log-likelihood = –2253.0)

Coefficient	Estimate	*SE*	$Pr(>\|z\|)$
Intercept	−0.93381	0.18449	< 0.0001
Dissimilation Faithful	2.56793	0.28062	< 0.0001
Dissimilation Second	−0.57989	0.11889	< 0.0001
Short-range	0.09676	0.16469	0.5568
Long-range	0.28368	0.16391	0.0835
S-Diss-M-Faith	0.29786	0.19036	0.1177
M-Diss-S-Faith	0.58095	0.18966	0.0022
Short-range × S-Diss-M-Faith	1.63310	0.23875	< 0.0001
Long-range × S-Diss-M-Faith	−0.40721	0.22436	0.0695
Short-range × M-Diss-S-Faith	−0.52090	0.22371	0.0199
Long-range × M-Diss-S-Faith	−0.92738	0.22379	< 0.0001

Table B.18: Summary of the fixed effects portion of the mixed-effects logistic regression for Experiment 4 with Long-range baseline (N = 4518; log-likelihood = –2253.0)

Coefficient	Estimate	*SE*	$Pr(>\|z\|)$
Intercept	−0.6561	0.1821	< 0.0001
Dissimilation Faithful	2.5685	0.2804	< 0.0001
Dissimilation Second	−0.5782	0.1189	< 0.0001
Short-range	−0.1817	0.1630	0.5570
Medium-range	−0.2757	0.1639	0.2648
S-Diss-M-Faith	−0.1034	0.1884	0.5832
M-Diss-S-Faith	−0.3421	0.1884	0.0694
Short-range × S-Diss-M-Faith	2.0337	0.2379	< 0.0001
Medium-range × S-Diss-M-Faith	0.3968	0.2243	0.0769
Short-range × M-Diss-S-Faith	0.4018	0.2221	0.0704
Medium-range × M-Diss-S-Faith	0.9179	0.2238	< 0.0001