UNIVERSITY OF CALIFORNIA

Los Angeles

A System for Morphophonological Learning and its Consequences for Language Change

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Linguistics

by

Dustin Andrew Bowers

2015

ABSTRACT OF THE DISSERTATION

A System for Morphophonological Learning and its Consequences for Language Change

by

Dustin Andrew Bowers

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 2015

Professor Bruce Hayes, Co-Chair

Professor Kie Ross Zuraw, Co-Chair

A major focus of linguistic research is characterizing adult knowledge of language and detailing how it is acquired. Language change, to the extent that it is driven by learners in response to observed adult data, is a valuable source of data for pursuing this topic. The shift from one language to another is only possible if the analytic preferences of language learners lead them to adopt a different analysis than that of their parents. A particularly noteworthy type of change is paradigm levelling, where some allomorphs of a morpheme are replaced by another allomorph.

This dissertation proposes a learning algorithm that replicates historically attested paradigm levellings. Previous attempts have restricted the inputs of phonological computation to be identical to a surface allomorph, so that paradigm levelling is triggered whenever a derived allomorph is not predictable from the base allomorph. Such a restriction is unnecessary. In the system proposed here, the absence of a grammar to explain the observed language is the trigger for levelling. In this case, the learner privileges the generation of a subset of a paradigm, and selects inputs that are appropriate to that task. The observed replacement of an allomorph by another in paradigm levelling is achieved by using an input that the grammar maps to the replacing allomorph, but which cannot be mapped to the replaced allomorph.

The learning algorithm proposed here makes accurate predictions for language change. Languages that are straightforwardly described in the assumed grammatical framework (parallel OT),

are diachronically stable. This stands in contrast to previous theories of levelling, which predicted diachronic instability for some paradigms in Russian. Furthermore, other languages are correctly predicted to undergo levelling, including a particularly dramatic case from Odawa. Ultimately, this helps substantiate the link between learning theory and language change.

The dissertation of Dustin Andrew Bowers is approved.

Donka Minkova

Ed Stabler

Kie Ross Zuraw, Committee Co-Chair

Bruce Hayes, Committee Co-Chair

University of California, Los Angeles

2015

To my family, who taught me fairness, happiness and love

and to Aaron, who left the world too soon

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Completing a dissertation is curiously chimerical, as it is both clearly attainable but consistently ephemeral. I would like to thank my advisers Bruce Hayes and Kie Zuraw for not only helping me confront the chimera but for cultivating all aspects of my development as an academic over these last five years. I hope to be as positive an influence for others as you have been for me. I also owe a great deal to my committee members Ed Stabler and Donka Minkova. Ed's exacting comments and precision can baffle me for months at a time, but the resulting crystalline understanding is incalculably valuable. Meetings with Donka have invariably given me a better notion of the true meanings of progress and success, and for that I am also grateful.

It can be difficult to adequately credit the communities and institutions that make our success possible. Hopefully naming them here will partially suffice. The faculty, graduate students and staff who make up the UCLA Linguistics department have collectively taught me how to practice competence and give only the best performance. I'd particularly like to thank my friends: Natasha Korotkova, Yu Tanaka, Adam Chong, Meaghan Fowlie, Joe Buffington and Vania Kapitonov for their companionship during this journey. Likewise, I would be remiss if I did not recognize my cohort members Yun Kim, Lauren Winans, Jun Yashima and Michael Lefkowitz. Further, none of this work could have been possible without my consultants, but especially without Reta Clement.

While I have often forgotten about it, there is a rich world outside of linguistics. I'd like to thank my dear friends Flint, Lily, Zarina, Alexa, Eva, Nick, Lisa, the two Rachels, Spike and Chris. Whether a denizen of Foster III or an assimilated satellite, we are all comrades of the quest. To the non-linguists in my LA circle, Jasper, Shanna and Katie, thank you for helping fill out this vibrant city for me. To the people who have known me longest, Andrew and Annette, Tim and Jenny, Keith and Carrie, thank you for being part of the village that raised me.

Finally, my far-flung and eclectically extended family deserves recognition for their support and patience through this adventure. Mom, Dad, Dina, Olin, Chelo, Alpen, Riven, Agemian, Delphi, Anand, Mackenzie, Asha and Kishor, I extend my heartfelt thanks for being at the core of my tribe. And to my wife, Kavita, thank you for holding me up; this dissertation is as much yours as it is mine. I can't wait to go do something even bigger with you.

<div align="center">VITA</div>

2007–2009	Tutor, Linguistics Department and Spanish Department, Reed College. Taught grammar and conversation in Spanish, core analytical techniques in Linguistics.
2009	B.A. (Linguistics), Reed College.
2011	Tutor, Linguistics Department, UCLA. Taught introductory topics to students in Linguistics 20.
2011–2012	Teaching Assistant, Linguistics Department, UCLA. Taught sections of Linguistics 20 (introductory linguistics), 161 (language documentation), 105 (morphology) under direction of Professors Ed Stabler and Russ Schuh.
2012–2013	Teaching Associate, Linguistics Department, UCLA. Taught sections of Linguistics 20 (introductory linguistics), 120A (introductory phonology), 165A (advanced phonology), under direction of Professors Martin Walkow, Robert Daland, and Kie Zuraw.
2014	Teaching Assistant Consultant and Summer Instructor, Linguistics Department, UCLA. Taught Linguistics 495 (Teaching Assistant training) and 120A (introductory phonology) under the direction of Professors Sun-Ah Jun and Bruce Hayes.
2012–present	Language Designer and Linguistic Consultant. Created languages and coached actors for films and television series.
2015 – present	Post-Doctoral Fellow, Linguistics Department, University of Alberta. Developed language technology for First Nations communities.

CHAPTER 1

Introduction

Since Kiparsky (1965) a major, if sometimes elusive, goal in generative phonology has been the elucidation of the link between synchronic grammar and language change. The premise is that principles of synchronic grammar delimit the space of possible analyses that learners can apply to data (see Kiparsky 1988; 1995; 2008). As successive generations of human learners are exposed to a language, the analyses they impose on the ambient language may be reflected in the changes the language goes through. The historical record thus potentially becomes like the log of a very long-running natural experiment. The challenge is to develop a nuanced and specific enough theory of grammar to allow the development of a language to be meaningfully discussed and explained. This dissertation, to corrupt a phrase from Jawaharlal Nehru, is an attempt to meet that challenge, not wholly or in full measure, but somewhat substantially.

Early discussions within this program were couched in rule-based phonology, and centered on questions like when two rules might change their relative order in the grammar, or when a rule might be lost altogether (see Kiparsky 1968b; 1971; 1973 and King 1969). In Optimality Theory (OT, Prince and Smolensky 1993 [2004]), more focus has instead been given to the significance of particular families of constraints for language change or the importance of architectural assumptions for the life cycle of phonological processes (Bermúdez-Otero 2006b; 2014a, Kiparsky 2008). However, in a loose parallel to the original discussions of rule re-ordering, some diachronically focused work in OT merely observes that different stages of a language can be characterized by different rankings of constraints.

The crux of the diachronic focus here is that a language may drift dangerously close to a system that is not generated by any ranking of constraints.[1] At such a point, the language must change; the

---

[1]By "dangerously close", I mean that the phonetic realization of a well-behaved phonological system may cross a

language must be restructured so that it falls into the typology defined by the ranking of constraints (see also Bermúdez-Otero and Hogg 2003 and Sanders 2003 for discussion of restructuring). The particular change that we focus on generating is paradigm levelling, where a single allomorph of a morpheme replaces other allomorphs over time. Creating a new system out of the ashes of an old system helps to fill in a lacuna left by Niyogi (2006), who demonstrated how a language community can gravitate to one analysis when multiple competing systems are already present.

In service of this goal, chapter 2 proposes a basic model of OT learning from overt surface forms. The goal of the learner is simply to identify whether or not there are rankings that generate the language it has been presented with, and if there are viable rankings, to show what they are.[2] If the learning model finds that the observed language is within the expected typology of languages defined by the constraint set, it does nothing to change the language or to value one viable analysis over another. The assumption is that any language within the expected typology is stable, or at least that if a language shifts from one type to another within the expected typology, that it is for reasons outside of the purview of our discussion.

A hoped-for advantage of such a bare bones learning model is that it so long as evaluation is parallel and constraints are given, it will arrive at an answer. However, this model is not completely neutral. In seeking all analyses of the data as permitted by the available constraints, it conflicts with theories that have a different conception of what a possible analysis is. Most salient among these theories is the Single Surface Base theory proposed by Albright (2002; 2005; 2010). Some time is spent in chapter 2 demonstrating that our model does indeed identify analyses that are not valid under that theory because our model makes the standard assumption that the underlying representation of a morpheme need not be identical to any one of its surface allomorphs.

The disparity between the Single Surface Base hypothesis and the model defended here is particularly important, as many of the arguments presented in favor of the Single Surface Base hypothesis rely on data from language change. Indeed, the Single Surface Base theory explicitly

threshold so that it no longer appears to be part of the typology defined by any ranking of constraints. See chapter 4 and Gess 2003, Bermúdez-Otero 2014a for further discussion of the role of phonetic drift and phonological reanalysis in language change.

[2]The simplicity of this goal puts the chapter in a fairly unique position with respect to prior literature. Though prior work in OT learning uses all of the methods and inference techniques that our model uses, to my knowledge this particular learning model has not been proposed.

claims that languages that require analyses from outside its hypothesis space are diachronically unstable. The ultimate development trajectory for such languages is that they are eventually reanalyzed by learners to comply with its more restricted hypothesis space.

The response to this is three-fold. First, chapter 3 surveys Russian paradigms where underlying representations of noun stems must contain obstruent voicing and vowel quality specifications, but both specifications never occur in the same allomorph. Our model predicts Russian to be diachronically stable, but the Single Surface Base theory predicts it to be diachronically unstable. Importantly, the predicted instability has not obviously resulted in a change for approximately 700 years. Second, chapter 4, which generally treats the reanalyses our model predicts, observes that some of the languages that have undergone reanalysis as predicted by the Single Surface Base theory are also predicted to undergo reanalysis by our model. The chapter also illustrates how the key intuitions of the Single Surface Base theory can be preserved within our model without sacrificing the descriptive advantages our model has over the Single Surface Base theory. That is, paradigm levelling can occur even if composite underlying forms are allowed. Finally, chapter 5 highlights the case of Odawa and other languages that developed rhythmic syncope. These languages undergo reanalysis, but nonetheless continue to require an analysis that is outside of the space assumed by the Single Surface Base theory. In contrast, the required analysis of the reanalyzed language is acheivable under our model.

Chapter 6 concludes our discussion. The primary conclusion is that the strong typological predictions of OT, which cannot generate many patterns previously handled with opaque rule orderings (see Bakovic 2007; 2011, McCarthy 2007b), are a useful starting point for investigations into what systems learners will and will not tolerate. The discussion shows in particular that reanalysis strikes, potentially quite dramatically, languages that have innovated patterns that OT cannot generate. This is far from an indictment against all opaque phonology; the vitality of opaque and transparent systems is an empirical question that is certainly not comprehensively addressed here. But it does highlight the utility of applying a synchronic system for morphophonological learning to diachronic changes.

CHAPTER 2

OT Learning

In its approach to learning phonology, this dissertation relies on concepts that have been extensively developed by prior work in the OT tradition (Tesar and Smolensky 1993; 1998; 2000, Tesar and Prince 2007, Merchant 2008, Jarosz 2006). To put it briefly, the learning device imagined here seeks all possible analyses of the data it encounters. In the context of phonotactic learning (Prince and Tesar 2004, Hayes 2004), it identifies every constraint that can dominate a loser-preferring markedness constraint (see section 2.3). In morphophonological learning, it adopts a perspective first proposed by Tesar and Prince (2007) and later elaborated on by Merchant (2008, see also Tesar 2013), where every permutation of unfaithfulness suggested by paradigmatic alternations is taken into consideration. The chief difference between this theory and prior work is that no attention is paid here to ranking biases or questions of efficiency, as the chief concern is simply the identification of analyses that are consistent with observed data. Chapter 4 discusses the most novel part of the learning proposal: specifying what happens when no analysis for the entire language is available. That discusion will rely crucially on the learner's ability to detect a failure, and on what the learner has nonetheless figured out about the language in such a case.

This chapter treats the identification of analyses that are consistent with observed data. Section 2.1 lays out basic assumptions and results from the Gold learning tradition (1967), and proposes some simple modifications to the assumptions in order to allow phonological learning to be treated in this tradition. Section 2.2.1 introduces the logical notation that underlies phonological learning. Sections 2.3 and 2.4 discuss how to use the information present in surface forms and alternations between paradigmatically related forms in learning. Section 2.5 illustrates the importance of allowing all paradigmatically related forms to contribute to URs.

2.1 Basic Gold Learning

A typical language learning algorithm in the Gold framework (Gold 1967, Blum and Blum 1975, Angluin 1980) provides a function $\phi()$ from corpora of positive examples to grammars. Using a slightly more technical notation, we describe the function as $\phi() : C \rightarrow G$, where C is the set of corpora and G is the set of grammars.[1] Results in learnability theory center on whether some class of languages can be learned (identified) in the limit by such a $\phi()$. A fairly elementary result is that any finite class of languages is learnable. The learning function $\phi()$ merely has to wait until the corpus has grown large enough to rule out all languages but one. The cardinality of the class of languages in OT is maximally the factorial of a finite number of constraints. Clearly, the class of languages defined by OT is finite, and hence learnable.

A key feature of OT makes realistic learning slightly different from the typical Gold learning problem. OT languages are not just collections of surface forms, but mappings from a rich base of inputs to surface forms. That is, an OT language is not a subset of Σ^*, but a total function from Σ^* to Σ^*. Recall that in the Gold paradigm, the function is defined on a corpus of positive examples. Given the definition of an OT language, a positive example is clearly an input-output pair. It is immediately obvious, however, that human language learners do not encounter input-output pairs, but rather encounter outputs alone. In other words, realistic data is impoverished beyond what is typically assumed.

This impoverishment of data has important repercussions. Most importantly, it is possible for OT grammars to generate the same sets of surface forms, even if they map the same inputs to different outputs. For example, with a constraint inventory like *VOIOBS#, ID-VOI and MAX (penalizing word-final voiced obstruents, the change of voice specifications or deletion, respectively), there are two languages that obey the ban against word-final voiced obstruents, depending on the relative rank of faithfulness constraints. In the first, all word-final voiced obstruents delete, as shown by the following tableau instantiating the ranking *VOIOBS# ≫ ID-VOI ≫ MAX:

---

[1]Though the learning function is defined on a corpus of data, this is not batch learning. Rather, the corpus at any particular time is a finite sample from an infinitely long text. More data can be drawn from the text, but there is never a guarantee that all data has been observed.

(1)

	ad	*VoiObs#	Id-voi	Max
a. ☞ a				*
b. ad		*!		
c. at			*!	

If the ranking is changed so that it is less costly to change voicing specifications than it is to delete segments, then the same input is mapped to a different output, as in the following tableau:

(2)

	ad	*VoiObs#	Max	Id-voi
a. ☞ at				*
b. ad		*!		
c. a			*!	

Because OT states that languages are functions from inputs to outputs, the two tableau above instantiate different languages, even though the surface forms of the languages are identical (i.e. the set of all strings not ending in voiced obstruents). Specifically, all words lack word-final voiced obstruents, though in all other environments either specification of voice in obstruents is permitted. A learner will therefore never encounter data that would disambiguate between two different languages if surface forms alone are considered. With data like this, it is clearly impossible for a learner to identify an arbitrary language in the class of languages defined by a set of OT constraints. The best that can be done is a learner that identifies sets of languages.[2] That is, the function must

---

[2]Concretely, this means that our learning function must map corpora into the powerset of the set of grammars, or in our technical notation: $\phi() : C \to \mathcal{P}(G)$. This has been informally proposed in the linguistics literature as least as early as King (1988). See section 2.2.1 below for a discussion of a compact representation of a set of OT grammars using Elementary Ranking Conditions (ERCs, Prince 2002).

produce from a corpus the set of grammars that is consistent with the corpus.

2.1.1 Alternations as a Window on URs

Real human languages do not consist of unanalyzed surface forms. This is because language data consists of sentences composed of words, and words are often composed of sub-parts that re-appear in other words. Put more succinctly, syntax and morphology exist. In keeping with the basic examples utilized so far, we will restrict our attention to word-level phonology, focusing on the form of morphemes and legal words. While proposals on the representation of morphological systems abound, the classic approach in generative phonology is that every instantiation of a morpheme is derived by the grammar from the same string (or, departing slightly from the classic approach, set of strings). The requirement that the surface realizations of a morpheme be derived from the same representations provides an important additional source of information for learning. Where individual surface forms are often compatible with a wide array of underlying representations, morphological analysis can reveal properties of URs.

Properties of the UR are revealed when a morpheme undergoes alternations, so that it does not take the same form in every instantiation. On each parameter (feature, epenthesis/deletion) where allomorphs differ, at most one allomorph can faithfully display the underlying specification. For instance, the regular English third person singular agreement morpheme has three allomorphs [-ɨz, -z, -s]. The UR for this morpheme can contain (or fail to contain) [ɨ], or contain a segment that is voiceless, voiced, or potentially underspecified for voice (Inkelas, Orgun and Zoll 1997). However, the UR cannot both contain [ɨ] and not contain it, or have multiple voice specifications.[3] The usual analysis proposed by linguists is that the plural morpheme is underlyingly /-z/, but from the perspective of a child who only has the allomorphs, it could be any of these combinations.

---

[3]The picture is slightly more complicated if morphemes have sets of URs, rather than a single UR. In order for a UR to be a member of the set for a morpheme, the grammar must map it to all allomorphs of the morpheme. Under this definition, a set of URs can contain URs with contradictory specifications, so long as the grammar derives the correct allomorphs forms from each UR in the set. In this case, it is still true that at most one feature of an allomorph is a faithful realization of the UR set, because no allomorph is a faithful realization of the feature as instantiated in every member of the set. UR sets that contain contradictory featural specifications differ meaningfully from underspecified URs because underspecification allows a three-way contrast between classes of, for instance, non-alternating voiceless segments, non-alternating voiced segments, and segments that alternate between voiced and voiceless realizations (see Inkelas, Orgun and Zoll 1997).

Alternations can provide information that disambiguates between grammars. For instance, consider a learner that is exposed to the language that prohibits word-final obstruent voicing, and the corpus contains related forms like [ad-a, at], so that one morpheme takes the allomorphs [ad, at]. The crucial information gained from the alternation is that one or both allomorphs is unfaithful to the underlying voice specification. As will be explored in some detail in section 2.4, the unfaithfulness signaled by the alternation has two repercussions for inferring a ranking. First, ID-VOI must be dominated by other faithfulness constraints, like MAX, so that the observed unfaithfulness is preferred over other unfaithful mappings. Second, ID-VOI must also be dominated by markedness constraints, in this case *VOIOBS# in order to force the observed unfaithfulness to be preferred over faithful mappings. The upshot of this is that the alternations provide data on what function the observed language belongs to, as the faithful map to be avoided specifies some aspect of the input, while the faithfulness constraint to be violated specifies what the input is mapped to. Within the context of the toy devoicing language, where there was formerly no way to distinguish between the two functions with the same output inventory, the allomorphy in [ad, at] is compatible with the feature changing function, but not the deletion function. As (3) shows, the feature changing ranking ensures that one allomorph is unfaithfully derived from some UR.

(3)

ad	*VOIOBS#	MAX	ID-VOI
a. ☞ at			*
b. ad	*!		
c. a		*!	

Still more importantly, the other allomorph can also be derived from the same UR with this ranking, as shown in (4). As mentioned above, the core requirement when analyzing alternations is that at most one allomorph can be derived faithfully on a particular parameter. This requirement has clearly not been violated by this analysis.

(4)

ad-a	*VOIOBS#	MAX	ID-VOI
a. ☞ ad-a			
b. at-a			*!

Meanwhile, a hierarchy like *VOIOBS# ≫ ID-VOI ≫ MAX specifies that it is better to delete a word-final voiced obstruent than it is to cause a voicing disparity. There is thus no way for a grammar of this sort to account for a paradigm with a voicing alternation. The language generated by the grammar does not contain the input-output mappings required by the alternations. Alternations thus provide another dimension that grammars can be tested for consistency with.

2.1.2   Paradigmatically Labeled Data

Given the utility of alternations in disambiguating between languages, we modify the basic Gold learning conception of learning data to include paradigmatic alternation. Every surface form is accompanied with a specification of the morphemes present in the form, and which sub-strings of the form are allomorphs of the morphemes. The traditional interlinear glossing format encapsulates such information, as in (5):

(5)   walk-s
      WALK-3.sg.pres

The information in (5) consists of an observed form, morpheme identifiers, and hyphens to indicate where the allomorphs of each morpheme are present. It is formally equivalent to a tuple consisting of an observed string, a tuple of the morphemes instantiated in the string, and a tuple of the indices where the allomorph of each morpheme occurs. For maximal clarity, a datum d is defined below:

(6) $d = \langle O, M, A \rangle$

9

Where O is an observed string, M is a tuple of unique morpheme identifiers m corresponding to the morphemes that appear in O, and A is a tuple a of tuples indicating where the allomorph of each morpheme is located in O. Thus, M and A are defined below.

(7) $\quad M = \langle m_n, m_{n+1} \ldots m_p \rangle$

(8) $\quad A = \langle a_n, a_{n+1} \ldots a_p \rangle$

The allomorph indices a in A are tuples of string positions, which indicate where in O the morphemes in M are realized. Counting in string positions starts from 1. Hence, the information in (5) is equivalently represented as $\langle \text{walks}, \langle \text{WALK}, 3.\text{sg.pres} \rangle, \langle \langle 1, 2, 3, 4 \rangle, \langle 5 \rangle \rangle \rangle$.[4] Since the traditional way of illustrating the morphological decomposition of a word is easy to read, I will typically use it over the equivalent notation developed here.

The utility of paradigmatic data is that it reveals unfaithfulness that must be derived by the grammar. To this end, it is useful to assume a function that identifies the disparities between allomorphs of a morpheme. For instance, if the learner is confronted with *slice-s* [slaɪs-ɨz], *slog-s* [slɑg-z] and *walk-s* [wɑk-s], the crucial information to be obtained from the allomorphy of the suffix is that in the first segment of the suffix there is an epenthesis/deletion disparity, while in the second segment there is a voicing disparity. It is possible to obtain such information mechanically using string alignment methods (see for instance, Ristad and Yianilos 1996 and Cotterell, Peng and Eisner 2014), or using finite state automata. The disparity information we will discuss is readily observed by visual inspection, so I will leave this unformalized.

2.1.3   Sophistication of Assumed Analysis

An analysis of words into their component morphemes and identifying their allomorphs is clearly quite sophisticated, but we will not concern ourselves with how it is achieved (see Goldsmith 2005, Golénia, Spiegler and Flach 2009 for computational perspectives, and Kim 2015 for an infant acquisition perspective). The crucial concern for our purposes is simply that the result of

---

[4]Note that this notation allows for discontinuous morphemes, like triliteral roots found in Semitic languages, but is awkward for morphological ablaut patterns or autosegmental morphemes. None of these issues will be crucial in the following discussion.

morphological analysis provide an indication of which strings come from the same underlying source.

2.1.4   Limits of Alternations

Note that even including paradigmatic information in the corpus is not necessarily enough to reveal all properties of inputs. This is true even if a limitlessly rich morphological inventory were supposed. The problem that arises is that high-ranking markedness constraints can forbid a marked structure in all contexts. Hence, even morphemes with the marked structure specified in the UR may never alternate, and certainly will not surface faithfully. If the ranking causes the allomorphs of the morphemes in question to display only a strict subset of the underlying features, there will not be direct evidence to force only the desired ranking to be available.

The simplest example of such a scenario comes in a language that permits voiceless obstruents, but not voiced obstruents. That such a language is possible is shown by the ranking *OBSVOI ≫ ID-VOI. Obviously, with an undominated ban against voiced obstruents, any morpheme that contains a voiced obstruent underlyingly will be devoiced. However, even though there is an unfaithful mapping being performed some of the time, nothing distinguishes the faithful surface forms from the unfaithful ones. With universal devoicing, there is no way to backtrack from paradigmatic information to the intended underlying specification.

The upshot of this is that paradigmatic data can be compatible with more than one OT grammar. Native speakers of a language often appear to have selected one grammar out of the set of compatible grammars. The most commonly cited evidence of this comes from phonotactic judgements, where for instance, observed [bɹɪk] and unobserved [blɪk] are accepted by native English speakers, but unobserved [*bnɪk] is not (Chomsky and Halle 1965, see also Daland et al 2011). The humans who have these judgements presumably have selected a grammar that does not just ensure observed forms are legal, but that some unobserved forms are legal while others are not. Crucially, this has been done without any additional information from alternations, since English lacks alternations that hinge on legal onset sequences.

A near-consensus view is that the final choice of grammar is maximally "restrictive", in the

sense that it permits the fewest number of unobserved surface forms to be legal. A variety of proposals for obtaining the most restrictive grammar have been put forth, including ranking biases (Prince and Tesar 2004, Hayes 1999b), seeking to maximize the probability of the observed data (Jarosz 2006, Hayes and Wilson 2008), or selecting underlying forms that are maximally different from their observed surface allomorphs (Tesar 2013 chapter 8). Our learning theory is not immediately concerned with the question of restrictiveness, and consequently will simply volunteer a set of grammars if the phonotactic data and alternation data are compatible with more than one ranking.

2.2 Logic of OT Ranking

The discussion thus far has focused on total rankings of OT constraints. While it is possible to simply enumerate every total ranking of a constraint set (Jarosz 2006), it is more common to incrementally accumulate partial ranking statements stated in the Elementary Ranking Condition (ERC) logic of Prince (2002). This section introduces ERCs and shows how they can be manipulated to determine the viability of a hypothesis.

2.2.1 Elementary Ranking Conditions

OT specifies that the surface forms of a language are the ones that optimally satisfy a total ordering (hierarchy) of constraints. Each constraint evaluates elements of $\Sigma^* \times \Sigma^*$ and assigns them a non-negative integer score indicating the degree to which the constraint "objects" to the element (0 expresses no objections, higher numbers express more severe objections). The preferences of constraints that are higher in the hierarchy take precedence over the constraints that are lower in the hierarchy. The elements that receive the lowest score from the highest ranked constraint are passed to the next constraint for evaluation, continuing down the hierarchy until only one element (the optimal element) remains. Rankings are typically written in prose as $\text{CON}_1 \gg \text{CON}_2$ (read CON_1 outranks CON_2, or CON_1 dominates CON_2). Ranking statements reflect a necessary condition for the observed winner to surface. In other words, there is a candidate which loses to the winner, and CON_1 prefers the winner to the loser, while CON_2 prefers the loser to the winner. In order for the

winner to win, the preference of CON_1 must take precedence over the preference of CON_2, which is reflected in stating $\text{CON}_1 \gg \text{CON}_2$.

An alternative to writing $\gg$ is to display the preferences of the constraints in a tableau, where W indicates winner-preference, L indicates loser-preference and e or whitespace indicates preference for neither. When constraint ranking requirements are displayed in this format, they are known as Elementary Ranking Conditions (ERCs).

(9)

	CON_1	CON_2
a. winner-loser	W	L

This representation compactly describes the available rankings that ensure that an intended winning candidate actually wins. Briefly, in every vector, there must be some W-valued constraint that outranks every one of the L-valued constraints. Note that a set of ERCs may compactly represent a large swath of total rankings, as multiple total rankings may respect the "some W over all L" requirement, thereby satisfying the ranking conditions. Thus, a set of ERCs is a natural representation for a set of grammars, which we argued in section 2.1 is the correct value for our learning function to take in response to learning data.

2.2.2 Inconsistency and Fusion

As ust indicated, much depends on there being a total ranking that ensures that in every ERC, there is some W-valued constraint that outranks the totality of L-valued constraints in that ERC. Informally, when that criterion is met, the set of ERCs under consideration represents a valid ranking of the constraints (and ultimately, the language data on which the ERCs are based can be generated by an OT grammar). Sets of ERCs that satisfy or fail to satisfy this condition are known as "consistent", or "inconsistent", respectively. Consistency can be decided using the widely known Recursive Constraint Demotion (RCD, Tesar and Smolensky 1993) or Fusional Reduction (FRed, Brasoveanu and Prince 2011) algorithms. This section briefly outlines how inconsistency is de-

tected using the fusion operation that lies at the core of both algorithms (Prince 2002).

In the simplest case of inconsistency, consider a set containing only a single ERC, where no constraint favors the intended winner and at least one constraint favors the loser. Since there is no constraint that voices a preference for the winner over the loser, the winner is harmonically bounded (see Samek-Lodovici and Prince Samek-Lodovici and Prince), and there is no ranking that ensures that the winner is generated and the loser is not. The ERC associated with a harmonically bounded winner, one containing L but no W is the value that signals inconsistency. In order to detect inconsistency when multiple ERCs are under consideration, the fusion operation is employed.

Fusion, denoted with the symbol $\circ$, produces an ERC that embodies the collective truth of a set of ERCs (cf Brasoveanu and Prince 2011). The fusion operation is defined on the values of the ERC logical system as shown in (10). In what follows, we will refer to the fusion of ERCs, which is merely the extension of the fusion operation coordinate-wise to multiple ERCs.

(10)　　$L \circ L = L$

　　　　$L \circ e = L$

　　　　$L \circ W = L$

　　　　$e \circ e = e$

　　　　$e \circ W = W$

　　　　$W \circ W = W$

Recall that the dominance relation is established in an ERC by the presence of the polar values W and L. As the fusion of a set of ERCs is meant to represent the collective ranking statements of the set, both L and W trump non-polar e in fusion. However, when it is necessary to decide between the polar values, L retains its dominant role, trumping W. This is because a loser-preferring constraint *must* be dominated, no matter if it may be a potential dominator in some other contest. When summarizing a set of ranking statements, wherever the strongest statement appears in the set, it must also appear in the summary.

Due to the dominant character of L, fusion detects inconsistent ERC sets. For instance, two ERCs that are individually consistent, but mutually inconsistent fuse to an inconsistent ERC in

(11).

(11)

	C_1	C_2
1.	W	L
2.	L	W
$1 \circ 2$	L	L

The first ERC in (11) requires that C_1 dominate C_2, while the second ERC requires that C_2 dominate C_1. Domination is not a reciprocal relation, so the truth of both statements cannot be accepted. Fusing the ERCs extracts this contradiction, showing that the ERCs, if taken together represent a ranking and hence a language that cannot be represented by OT. If any subset of a set of ERCs fuses to a vector with L, but not W, inconsistency results. In what follows, we will simply identify inconsistent subsets visually, though the Recursive Constraint Demotion algorithm has been proven to find any inconsistency (Tesar and Smolensky 1993), and while the Fusional Reduction algorithm has only a forthcoming correctness proof (Prince and Brasoveanu 2010), at the time of this writing, a case where it fails to identify an inconsistent set of ERCs has not been reported.

2.2.2.1 Fusion and Consistency

In what follows, we will sometimes use fusion on consistent sets of ERCs, as fusion produces an ERC that is entailed by the joint truth of the fused ERCs (see Brasoveanu and Prince 2011). This allows familiar chains of reasoning about constraint ranking to be formally produced. For instance, if C_1 dominates C_2, and C_2 dominates C_3, then we can conclude by transitivity of domination that C_1 also dominates C_3. (12) demonstrates how fusion produces the ERC corresponding to this conclusion.

(12)

	C_1	C_2	C_3
1.	W	L	e
2.	e	W	L
$1 \circ 2$	W	L	L

As a more challenging example, the same conclusion as that reached in (12) can be reached from a weakened version of the original ERCs. If we accept that either C_1 or C_3 may dominate C_2, and we also accept that C_1 or C_2 may dominate C_3, then we accept that C_1 must dominate both C_2 and C_3. The first ERC establishes that C_2 must be dominated, while the second ERC establishes that C_3 must be dominated, and C_1 is the only available constraint to dominate both of them. This is illustrated in (13).

(13)

	C_1	C_2	C_3
1.	W	L	W
2.	W	W	L
$1 \circ 2$	W	L	L

2.3 Dominated Markedness

Having established ERCs as the representations of constraint ranking, we turn now to how ERCs may be collected from data during learning. Following the traditional division of phonological learning into phonotactic learning and morphophonological learning, this section treats the collection of ERCs from surface forms alone, while section 2.4 discusses the collection of ERCs while attending to alternations.

As mentioned at the beginning of this chapter, the goal of this learning model is simply to identify all possible analyses latent in the constraint inventory for the observed data. In the context of phonotactic learning, such a "leave-no-stone-unturned" approach is concerned only with finding the set of all grammars under which the observed surface forms are legal.[5] In other words, within

---

[5]Readers should not be surprised to find no discussion of efficiency or restrictiveness, which often dominates discussions of phonotactic learning (see Prince and Tesar 2004, Hayes 2004, Jarosz 2006, Hayes and Wilson 2008).

this restricted domain of phonological learning, the goal is to find the set of grammars such that any observed surface form can be the optimal output of some input. What must be avoided are grammars under which an observed form is not a viable output of any input.

The grammars that cannot generate observed surface forms share a central characteristic. Some markedness constraint that is violated by a form in the corpus fails to be dominated by a constraint that favors the form. Calculating which markedness constraints are violated by forms in the corpus is straightforward (the markedness constraints are simply applied to each form in the corpus). It is less straightforward to calculate which constraints are available to dominate the violated markedness constraints. This is primarily because some of the constraints that could favor the observed form are faithfulness constraints, and the value of the UR is not given with the surface form.

The classic approach to this problem (proposed independently by Prince and Tesar 2004 and Hayes 2004) starts from the observation that marked structures in the corpus are either the result of high-ranking markedness or faithfulness constraints. In the former case, all surface forms must obey the restriction enforced by the high-ranking markedness constraint, which comes at the cost of the violated markedness constraint. In the latter case, the markedness violation is potentially due to idiosyncratic properties of the word in question; that is, the marked sequence is the faithful realization of material specified in the UR. Any algorithm is remiss if it fails to represent possibilities, so the classic approach assumes that the UR for every surface form in the corpus is identical to that surface form. In the basic case (though see 2.3.3), this identity assumption maximizes the opportunities for faithfulness constraints to be the cause of marked structures in the corpus, so no possibility is missed. In the next section, we consider more carefully the importance of the identity assumption.

2.3.1 Utility of Identity URs

Recall that the goal of phonotactic learning is find the markedness constraints that surface forms violate, and identify all constraints that can dominate them. By temporarily setting the UR to be identical to the observed SR, the classic approach maximizes the number of constraints that favor the the observed SR. The basic reasoning behind this is straightforward. First, faithfulness

constraints enforce similarity between the UR and the SR, and so the identity UR ensures that faithfulness constraints will never disfavor the observed SR. Only markedness constraints remain to be discussed. Unlike faithfulness constraints, markedness constraints are blind to URs, and thus their evaluation of SRs is constant. With no upper bound on the number of violations a markedness constraint can assign, the maximum number of markedness constraints that can favor an SR is the total number of markedness constraints. No matter how many violations a candidate output incurs on a constraint, there is always another candidate output that would incur more violations. It is better instead to ensure that all of the markedness constraints that must favor the intended SR have been found.

Work by Riggle (2004) emphasizes that it is irrelevant to be concerned with all candidate outputs, but that focus should instead center on the non-harmonically bounded outputs. To this end, Riggle proposes the *contenders()* algorithm to compute all of the non-harmonically bounded outputs from an input. The sets produced by the algorithm have the following characteristics. First, unless the set of faithfulness constraints is degenerate, the fully faithful output is not harmonically bounded and is included in the set (see below). Second, all other members of the set, if they exist, must outperform the fully faithful output on some markedness constraint c_1, since otherwise they will be harmonically bounded by the fully faithful output. Third, any markedness constraint c_2 that favors the fully faithful candidate to a non-faithful candidate must do so as the minimal cost of the better performance obtained on c_1 (otherwise some other non-faithful candidate will harmonically bound it). This final characteristic forms the basis of the guarantee that only markedness constraints that necessarily favor the fully faithful output will be found. Because all non-harmonically bounded candidates are located by the *contenders()* algorithm, we know that all such constraints will be located. Because these markedness profiles are independent of faithfulness constraints, we know that the assumption of the identity UR does not interfere with locating them.

The assumption of the identity UR has an important final consequence. Recall from the previous paragraph that, except in degenerate constraint inventories that lack some crucial faithfulness constraint, the fully faithful candidate is never harmonically bounded. The reason for this is clear: by definition, faithfulness constraints never assign violations to the fully faithful candidate, but they do assign violations to unfaithful candidates. At the very least, the fully faithful candidate

will always be favored to an unfaithful candidate by faithfulness constraints, and so will not be harmonically bounded. This is important because it ensures that the intended winner will always be a contender (possible winner), and so it will always be possible to construct ERCs pitting the intended winner against the unfaithful contenders. Finally, sets of ERCs collected with the identity UR can never be inconsistent, because faithfulness constraints uniformly prefer the intended winner, ensuring that there is a class of constraints that places W and not L in every ERC.[6]

2.3.2 Worked Example

To see how this works, consider a toy language that permits voiced obstruents everywhere except word-finally, and permits voiceless obstruents in all environments.[7] A representative sample of the lexicon is shown in the table below.

(14)

Nominative	Accusative	Gloss
ʌk	ʌk-ʌ	'oar'
ʌk	ʌg-ʌ	'paddle'

Following the lead set by Tesar and Prince (2007), assume that this language is part of the typology defined by the following constraints:[8]

(15)   a.   *VOIOBS: assign one violation for every voiced obstruent in an output.

       b.   *VOIOBS#: assign one violation for every word-final voiced obstruent in an output.

       c.   *IVV: assign one violation for every intervocalic voiceless obstruent in an output.

       d.   MAX: assign one violation for every segment in the input that has no correspondent in the output.

       e.   DEP: assign one violation for every segment in the output that has no correspondent in the input.

---

[6]It is this fact that underlies Riggle's 2006b observation that the "identity grammar is a perennially available hypothesis".

[7]This language forms a running example to which I will add as need arises.

[8]This constraint set departs slightly from Tesar and Prince (2007) in including constraints that penalize insertion and deletion. Tesar and Prince minimize discussion of insertion or deletion operations because they make assigning correspondences more difficult.

f. ID-VOI: assign one violation for every output value of VOICE that differs from the corresponding input value.

Two of the words in the table above (*ʌk-ʌ* 'oar-acc, and *ʌg-ʌ* 'paddle-acc'), violate markedness constraints. Since we are only concerned with the domination of markedness constraints, we do not need to be concerned with information from alternations, which suggest faithfulness constraints to be dominated. Morpheme boundaries can in fact be ignored during this section.[9] These markedness constraints must be dominated for the forms to be legal. In the case of *ʌkʌ* 'oar-acc', the constraint banning intervocalic voiceless obstruents must be dominated. Taking the identity UR, as in (16) allows us to see what markedness constraints and what faithfulness constraints can ensure this result.[10]

(16)

ʌkʌ	*OBSVOI	*IVV	*OBSVOI#	MAX	DEP	ID-VOI
a. ☞ ʌkʌ		*				
b. ʌʌ, ʌk, kʌ		L		* W		
c. ʌgʌ	*W	L				*W

The losing candidates all avoid intervocalic voiceless obstruents, whether by deleting a segment (candidates b), or changing the voice specification of the obstruent (candidate c). The deletion candidates only incur a faithfulness violation, leaving only one possibility for the ranking of *IVV and MAX. Candidate (c) on the other hand, violates a potential ban against voiced obstruents, introducing a disjunction in the available grammars, so that *IVV is dominated by ID-VOI or by *OBSVOI.

---

[9]The situation would obviously be different if our markedness constraints referred to morpheme boundaries. See Martin (2007).

[10]Note that learning from unanalyzed surface forms is typically carried out by applying a ranking algorithm, perhaps supplied with a bias, to a set of ERCs (Prince and Tesar2004, Tesar and Prince 2007, Tesar 2013). The resultant ranking of constraints is used until a non-faithful loser wins, at which point the ERC from the most recent evaluation is added to the ERC set, and a new total ranking is constructed. We have adopted the more direct approach advocated by Riggle (2004 chapter 5), where the constraint set is kept unranked, but the full set of contenders is calculated, and the ERC associated with each loser is collected.

The other relevant surface form for markedness domination is ʌgʌ 'paddle-acc', which violates the constraint against voiced obstruents. From the identity UR, this can be resolved either by deletion of the obstruent or by devoicing it:

(17)

ʌgʌ	*OBSVOI	*IVV	*OBSVOI#	MAX	DEP	ID-VOI
a. ☞ ʌgʌ	*					
b. ʌʌ	L			*W		
c. ʌkʌ	L	*W				*W

The devoicing candidate is the most noteworthy at this point, because the markedness constraints that assign L and W assigned the opposite values when we considered the form ʌkʌ́ 'oar-acc'. One potential explanation for the presence of a voiceless intervocalic obstruent was that voiced obstruents were banned, while one potential explanation for the voiced obstruent is that voiceless intervocalic obstruents are banned. If either explanation were true, observed surface forms of the language would be illegal, so both must be false. In order to formally express this intuition, it is necessary to union the two sets of ERCs that have been collected thus far. This is not a dangerous move, because, as mentioned above, ERCs collected exclusively from identity URs are guaranteed to be consistent, since the entire class of faithfulness constraints will never assess L to an observed form. The unioned sets are presented below:

(18)

*OBSVOI	*IVV	*OBSVOI#	MAX	DEP	ID-VOI	Comparison
L			W			ʌgʌ → ʌgʌ, *ʌʌ (17)
L	W				W	ʌgʌ → ʌgʌ, *ʌkʌ (17)
	L		W			ʌkʌ → ʌkʌ, *ʌʌ (16)
W	L				W	ʌkʌ → ʌkʌ, *ʌgʌ (16)

As it is currently displayed, the set of ERCs in (18) does not clearly state the intuitive conclusion from the last paragraph, as lines 2 and 4 merely maintain the contradictory disjunctions. What

is desired is a set of ERCs that sums up our intuition that neither *IVV nor *OBSVOI necessarily outranks the other, but that ID-VOI outranks both of them. The FRed algorithm produces this conclusion, along with the conclusion that MAX must outrank both *OBSVOI and *IVV (derived from lines 1 and 3 in the ERC set in (18).

(19)

*OBSVOI	*IVV	*OBSVOI#	MAX	DEP	ID-VOI	Source
L	L		W			ʌgʌ → ʌgʌ, ʌkʌ → ʌkʌ, *ʌʌ
L	L				W	ʌgʌ → ʌgʌ, ʌkʌ → ʌkʌ, *ʌʌ

The table in (19) summarizes what has been learned so far about the rankings. In the first row, we see the conclusion that voiced obstruents and intervocalic voiceless obstruents must not be deleted, while the second row shows that they cannot be changed from their underlying specifications for voice. All of the unfaithful mappings that could allow a marked structure to be avoided are not available, since the marked structure must be generated. The basic pattern of a voicing contrast in onsets has been acquired, while no decision has been made on voice in codas.

2.3.3 Beyond the Basic Case

A critical, though for our purposes ultimately tangential, issue is the correctness of our learning strategy for all constraint types. As discussed in 2.3.1, ensuring the legality of a corpus of surface forms by assuming identity URs is a strategy that will always produce a feasible grammar so long as standard markedness and faithfulness constraints are used. Most importantly, all potentially viable grammars are represented at any particular point. The reason for this is that only non-feasible grammars are weeded out in such a system (for a technical discussion of what characterizes these constraints, see Tesar 2013). However, when the constraint set is expanded beyond these core constraints, assuming only identity URs means that the full range of feasible grammars is no longer represented.

As an example of a language whose proper analysis cannot be reached if identity URs are assumed, consider Extarri Navarese Basque (Kirchner 1995 and references therein). As shown

in (20), underlying /a/ is realized as [e] when preceding a vowel, while in the same environment, underlying /e/ is realized as [i], and underlying /i/ as an extra high [i^j].

(20)

Indefinite	Definite	Gloss
alaba bat	alabe-a	'daughter'
neska bat	neske-a	'girl'
seme bat	semi-e	'son'
ate bat	ati-e	'door'
erri bet	erri^j-e	'village'
ari bet	ari^j-e	'thread'

To generate such a system, the grammar must permit small moves along the phonological scale (as in /e/ → [i]), but not large ones (as in /a/ → [i]). This is easily translatable into constraint ranking: a faithfulness constraint banning large phonological jumps must outrank the constraint that favors one end of the phonological scale, which outranks a faithfulness constraint banning small phonological disparities. The problem that emerges in learning this with identity URs is straightforward: an intermediate value on the scale is mapped to itself. In setting the grammar so that /e/ →/ [e] (and not the less marked competitor [i]), the short-distance faithfulness constraint must be ranked over the markedness constraint. That is, the target analysis of /e/ → [i] is explicitly precluded as a result of the identity UR assumption.

Note that this failure is not due to having an incorrect constraint inventory. Learning with identity URs simply rules out potentially valid analyses if a correct constraint inventory is assumed. When a restricted phonological inventory of y, z out of possible structures x, y, z presented, a chainshift where $x \to y$ and $y \to z$ but $x \nrightarrow z$ is not identified as a potential analysis. This can be easily seen by considering a simplified version of the Basque case, where [e] and [i] are legal surface segments in hiatus, but not *[a]. Even if we assume a constraint set that can generate a chain-shift under a particular ranking (as in 21), such a grammar is ruled out by the phonotactic information alone.

(21)   a. *NON-HI/_V: Assign one violation mark for a segment that is [-HIGH] or [+LOW]

23

preceding a vowel. Assign two violation marks for a segment that is $\begin{bmatrix} -\text{HIGH} \\ +\text{LOW} \end{bmatrix}$ and preceds a vowel.

b. ID-HI: Assign one violation mark for a segment whose output value of [HIGH] differs from its input value.

c. ID-LOW: Assign one violation mark for a segment whose output value of [LOW] differs from its input value.

d. ID-HI & ID-LOW: Assign one violation mark for a segment whose output values of both [HIGH] and [LOW] differ from their input values.

The constraint inventory in (21) contains the necessary ingredients not only to describe a surface inventory with legal [i, e] and illegal *[a] before vowels, but to do so with a chainshift. The markedness constraint *NON-HI sets up a phonological scale with one end (low vowels) more marked than mid vowels, which are worse than the least marked high vowel category. Faithfulness constraints penalizing small moves along the scale are present (21b-21c), along with a conjoined constraint that bans long moves along the scale (21d). Importantly, encountering [e] provides evidence that ID-HI outranks *NON-HI:

(22)

ee	*NON-HI/_V	ID-HI	ID-LOW	ID-HI & ID-LOW
a. ☞ ee	*			
b. ie	L	*W		

Upon encountering this positive example, a learner using identity URs has precluded the chainshift analysis where /ee/ → [ie]. The hallmark of the chainshift analysis is that the conjoined faithfulness constraint outranks the markedness constraint, which outranks the constraints regulating small phonological disparities. It is this last part that has been placed out of reach by the identity UR assumption.[11] Even though no ERC places restrictions on the conjoined constraint which

---

[11]This is not to say that a grammar generating only [e, i] and not *[a] before vowels is out of reach. The ranking

24

makes chainshifts possible, chainshifts are still impossible.

Finally, this is clearly not a problem that more data can remedy. Phonological learning in OT seeks a consistent ranking, and the ERC obtained in (22) simply is inconsistent with the ranking that generates a chainshift. The best that could happen if alternations suggested a chainshift is inconsistency.

2.3.3.1 Chainshifts are not Unlearnable

The failure of one learning strategy does not mean that chainshifts are unlearnable. The failure of the identity UR strategy is straightforwardly attributable to the choice of a particular UR for deducing rankings. Choosing URs that are consistent with a chainshift would clearly allow the ranking that drives a chainshift to emerge. The target ranking to generate a chainshift is ID-HI & ID-LOW ≫ *NON-HI/_V ≫ ID-HI, ID-LOW. For instance, if /ee/ were allowed as a source for [ie], *NON-HI must outrank ID-HIGH, as is desired in a chainshift:

(23)

ee	*NON-HI/_V	ID-HI	ID-LOW	ID-HI & ID-LOW
a. ☞ ie		*		
b. ee	* W	L		

Furthermore, the ranking requirements in (23) are compatible with a grammar that maps /ae/ to [ee]. If a learner is freed from considering only identity URs, a working grammar that was precluded re-emerges as a potential analysis.

_____

could still be refined so that /ae/ → [ee], /ee/ → [ee], and /ie/ → [ie].

(24)

	ae	*NON-HI/_V	ID-HI	ID-LOW	ID-HI & ID-LOW
a. ☞	ee	*		*	
b.	ae	** W	L		
c.	ie	L	* W		* W

Abandoning the assumption of identity URs presents some logistical challenges. Clearly, once unfaithful URs are posited, there is no longer a guarantee that all faithfulness constraints will favor the intended winner, making it so that it is no longer safe to pool all the ERCs obtained from observed words into the same set. Rather, much as in morphophonological learning (see section 2.4), several sets of ERCs must be maintained, each corresponding to a combination of assumed URs. Of course, a brute force enumeration of the full space of possible inputs is an infeasible task. To get around this, one might propose that the full set of total rankings is enumerated, and each ranking made into a transducer following the algorithm in Riggle (2004 chapter 7). From these transducers, it is straightforward to see whether an observed surface form is legal in the grammar defined by the total ranking, and if it is legal, what the entire space of inputs that map to the observed surface form is. To my knowledge, there is currently no algorithm that could avoid enumerating the entire set of total rankings and determine the viability of unfaithful URs.[12]

None of the phonological systems considered in the remainder of this work require analyses that would be precluded by the identity UR assumption. The traditional learning approaches used thus far will be valid for our purposes. For further discussion of the issues involved in chainshifts, see Magri (2015a; Magri).

---

[12]Such an algorithm would presumably work much like the *contenders()* algorithm, which finds the winners that would be generated by any ranking of the constraints, except rather than find the non-harmonically bounded outputs for an input, it would find all of the inputs for which an output is not harmonically bounded.

2.4 Dominated Faithfulness

The above discussion makes it clear that phonotactic learning, which is concerned with the legality of a corpus of surface forms, depends on markedness constraints being dominated. In some cases, the markedness constraint can be dominated by another markedness constraint or a faithfulness constraint, while in others only faithfulness constraints favor the winner. In contrast, the legality of a corpus of paradigms relies on faithfulness constraints being dominated in order to explain any observed alternations. In order to completely describe a corpus of paradigmatically related forms, a learner must find a ranking that generates the observed alternations without rendering any observed surface forms illegal. Elaborating on Tesar and Prince 2007, this is carried out by collecting ERCs from unfaithful URs, and checking them for consistency against the set of ERCs collected from identity URs. Merchant 2008 and Tesar 2013) both propose learning strategies that differ from this one only in detail. This section discusses the analyses that are suggested by paradigmatic alternations and illustrates the exploration of those options.

Paradigmatic alternations do not by themselves reveal what the underlying form of a morpheme is. Rather, the differences between allomorphs of a morpheme indicate the features that the grammar causes to alternate. As discussed earlier, a maximum of one surface variant can be faithful to the underlying representation, but there is no indication which allomorph, if any, is faithful.[13] In such a situation, a viable strategy is simply to exhaust all of the options suggested by the alternations.[14] This can be easiest to see with a worked example. Recall the toy language introduced in the last section, whose current lexicon is reproduced below:

(25)

Nominative	Accusative	Gloss
ʌk	ʌk-ʌ	'oar'
ʌk	ʌg-ʌ	'paddle'

The paradigm for 'paddle' has a voice alternation, as can be clearly seen by opposing the

---

[13] Albright (2002; 2005; 2010) has proposed that children assume that one allomorph of a morpheme faithfully reflects the UR, see the review in section 4.6.

[14] Tesar (2013) proposes that candidate inputs form a complete atomic Boolean algebra, which negates the need for exhaustive search. Magri (2015b) shows that this property does not hold if many phonological features commonly used in phonology are used.

nominative stem allomorph ʌk with the accusative stem allomorph ʌg. There is no way to tell which allomorph, if any, is faithful simply by considering these allomorphs by themselves. The only available inference is that at most one of them is faithful. Hence, at the very least, potential URs for this paradigm are /ʌk/ and /ʌg/, as each of them differs from all allomorphs but one.

The next step is to gather the ERC sets generated by actually using each UR to generate the paradigm in question. That is, we must see what rankings must hold for /ʌk/, when embedded in affixes as necessary, to generate [ʌk, ʌg-ʌ], and likewise for /ʌg/. No rankings must be instantiated for /ʌk/ to map to [ʌk], because no constraints are violated by the mapping (causing all other candidates to be harmonically bounded).

(26)

ʌk	*ObsVoi	*IVV	*ObsVoi#	Max	Dep	Id-voi
a. ☞ ʌk						

In generating the accusative form [ʌg-ʌ] from /ʌk-ʌ/, however, faithfulness has been violated, and so must be dominated.

(27)

ʌk-ʌ	*ObsVoi	*IVV	*ObsVoi#	Max	Dep	Id-voi
a. ☞ ʌgʌ	*					*
b. ʌkʌ	L	*W				L
c. ʌʌ, kʌ, ʌk	L			*W		L

The collective ERCs from these paradigmatic maps are a morphophonological ERC set. They represent the required rankings that would generate the paradigm for 'paddle' in our toy language. The goal, however, is to find a ranking that both ensures all observed surface forms are legal and generates all paradigms. Because the phonotactic ERC set encapsulates all grammars that allow

28

observed surface forms to be legal, it can serve as a baseline against which to test morphophonological ERC sets. If a morphophonological ERC is inconsistent with the phonotactic ERCs, the morphophonological ERC set must require a ranking that would cause some observed surface forms to be illegal. This is the case with the current morphophonological ERC set, as can be seen from the union of the morphophonological ERC set with the phonotactic ERCs:

(28)

*OBSVOI	*IVV	*OBSVOI#	MAX	DEP	ID-VOI	Source
L	L		W			Phonotactics (19) ʌgʌ → ʌgʌ, ʌkʌ → ʌkʌ, *ʌʌ
L	L				W	Phonotactics (19) ʌgʌ → ʌgʌ ʌkʌ → ʌkʌ, *ʌkʌ, ʌgʌ
L	W				L	Alternation /ʌk-ʌ/ → ʌgʌ, *ʌkʌ
L			W		L	Alternation /ʌk-ʌ/ → ʌgʌ, *ʌʌ

However, this set of ERCs is inconsistent, as can be seen by fusing the second and third lines, which produces an ERC that only contains "L". There is no ranking that can satisfy the requirements represented in this ERC set. Since the phonotactic ERCs rule out all grammars that make observed surface forms illegal, the fault must lie in the morphophonological ERC set. The /ʌk/ hypothesis must be abandoned. Fortunately, the learner can still consider /ʌg/. The following tableau shows that /ʌg/ can indeed map to [ʌk].

(29)

/ʌg/	*OBSVOI	*IVV	*OBSVOI#	MAX	DEP	ID-VOI
a. ☞ ʌk						*
b. ʌg	*W		*W			L
c. ʌ				*W		L
d. ʌgʌ	*W				*W	L
e. ʌkʌ		*W			*W	L

When collecting phonotactic ERCs, the conditions for mapping /ʌgʌ/ to [ʌgʌ] were already collected, and they will not be repeated here. Importantly, the most recent morphophonological ERC set is consistent with the phonotactic ERC set. Below is the union of the phonotactic ERC set with the /ʌg/ ERC set.

(30)

*OBSVOI	*IVV	*OBSVOI#	MAX	DEP	ID-VOI	Source
L	L		W			Phonotactics (19)
L	L				W	Phonotactics (19)
W		W			L	Alternations /ʌg/ → ʌk, *ʌg
			W		L	Alternations /ʌg/ → ʌk, *ʌ
W				W	L	Alternations /ʌg/ → ʌk, *ʌgʌ
	W			W	L	Alternations /ʌg/ → ʌk, *ʌkʌ

Summarizing the set of ERCs in (30) in the notation of constraint domination produces the following constraint hierarchy: MAX, DEP, *OBSVOI# ≫ ID-VOI ≫ *OBSVOI, *IVV. At this point, all available information has been extracted from the observed paradigms and surface forms.

30

The grammar enforces a ban against final voiced obstruents at the cost of faithfulness to voice, while in other environments both voiced and voiceless obstruents are allowed. A total ranking of constraints has not been achieved, but out of an initial set of $6! = 720$ available rankings, the total number of available rankings has shrunk to $3! \times 2! = 12$.[15] Even though the learner has not identified a unique language from the class of OT languages, the set of languages that it has identified have all the desired paradigmatic properties. At this point the learner has converged. To see a summary of the procedure used in pseudocode, please consult section 2.6.

2.5 Composite URs

The discussion in section 2.4 suppressed a potentially controversial point. Individual features or segments of allomorphs in our model may be unfaithful to an underlying specification. This is opposed to an alternative view where the UR must be an observed allomorph, so that there is always an allomorph that is faithful to the underlying specification. Under the view defended here, the UR of a morpheme may differ from all of its observed allomorphs. This distinction has a long history in phonological theory, traceable at least to Kenstowicz and Kisseberth (1977 chapter 2) and appearing more recently in discussions of exemplar models or Albright's (2002; 2005; 2010) arguments in favor of the Single Surface Base hypothesis (see section 4.6). This section briefly outlines the various degrees of closeness that URs can theoretically have to surface allomorphs, and illustrates how the system outlined here can acquire a lexicon of URs that differ from all observed allomorphs.

In classifying theoretical positions on the abstractness of underlying forms, we might distinguish theories by the amount of leeway given to the learner in determining the UR. The most restrictive position, advanced most recently by Albright (2002; 2005; 2010), is that the UR for a morpheme must be a surface allomorph, and the URs for all morphemes of a particular category must be the allomorph that appears in a particular morphological or phonological environment. Slightly less restrictive is the theory that maintains that the UR for a morpheme must be a surface allomorph, but URs need not be drawn exclusively from the same morphological or phonological

---

[15]Though this is not generally true, in this case all twelve remaining rankings define the same input-output mapping.

environment. Less restrictive still is the model used here, the learner might be permitted to construct URs where input segments that alternate in surface forms may contain any of the features or segments observed in the correspondents of those segments in any allomorph. We might call such URs 'composite URs', and it is a widespread, though not universal, practice in OT phonology to assume URs at this level of abstraction.[16] At the most abstract, underlying representations might not be required to have any features in common with their surface realizations. As mentioned at the end of section 2.3.3.1, learning in such a theory is not completely hopeless. However, such analyses are often charged with being proposed simply to make the synchronic analysis of a language recapitulate its diachronic development. Interestingly, Kiparsky (1968a; 1973) argues that historical changes do not progress according to the predictions of an abstract UR model, though the issue was never fully settled, since there are longstanding patterns where an abstract UR analysis is attractive (Piggott 1971).[17]

To see how our model constructs composite URs, suppose the toy language is enriched with two further phonological regularities: word-final stress and adherence to a ban on unstressed [ɔ]. The language now appears as below:

(31)	Nominative	Accusative	Gloss
	ʌ́k	ʌk-ʌ́	'oar'
	ʌ́k	ʌg-ʌ́	'paddle'
	ɔ́k	ʌk-ʌ́	'rudder'
	ɔ́k	ʌg-ʌ́	'mast'

With the addition of these new regularities, the observed paradigms have alternations in multiple features. Most importantly, recombining the features that appear in the surface allomorphs results in URs that correspond to no observed allomorph. This can be seen most clearly in the paradigm for 'mast', which contains a nominative form ɔ́k and an accusative ʌg-ʌ́. With stress,

---

[16] 'Heteroclitic UR' has also been suggested as a name, but this appears to conflict with a pre-existing technical definition (Stump 2006).

[17] An alternative way of classifying analyses is not by the degree of abstraction in the URs, but in the amount of forms that must be labeled as irregular. An analysis that uses concrete URs will often require many forms to be analyzed as irregulars, while an analysis that uses highly abstract URs will tend to treat more forms as regular. A composite UR analysis, being intermediate in terms of concreteness, is similarly intermediate in terms of regularity.

vowel rounding, and consonant voice alternating, there are three binary features whose values can be recombined to make URs. Importantly, only URs the /ɔ́g/ or /ɔg/ can be mapped to the observed allomorphs by a grammar that allows all surface forms to be legal. These two URs are both composite URsThe space of URs suggested by the alternations of this morpheme is shown in Figure 2.5.
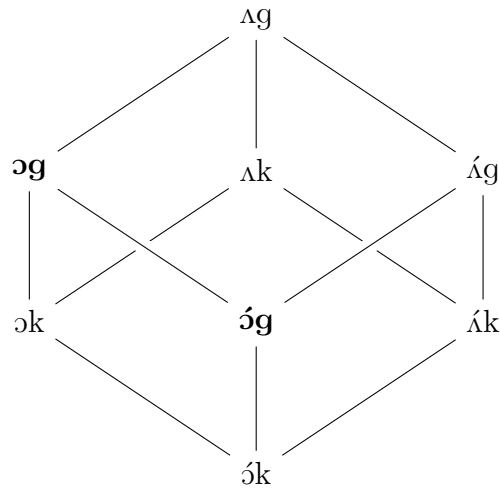
Figure 2.1: Hasse diagram of composite and concrete URs for the paradigm containing ɔ́k and ʌg-ʌ́. The composite URs that can support the observed alternations are bolded.

To informally substantiate the claim that only the bolded URs in Figure 2.5 can map to both ɔ́k 'mast' and ʌg-ʌ́ 'mast-acc', recall the intended phonological analysis of the enriched language. As before, there is a general contrast in obstruent voicing except word-finally, and vowel rounding is contrastive in stressed, but not unstressed syllables. In a system like this, obstruents in word-final position and vowels in unstressed syllables may be the surface realization of multiple underlying specifications. Of course, vowels in stressed syllables and obstruents in non-word-final position cannot be neutralized and so can only be the surface reflex of a single underlying specification. In seeking possible URs to link ɔ́k and ʌg-ʌ́, each form in the paradigm has a different contrastive element to contribute to the UR.

A more formal demonstration of the analysis first requires an expansion of the constraint set to describe the new phenomena. These constraints are shown below.

(32)   a.   *ɔ : Assign one violation for every unstressed [ɔ] in an output.

b. *ʌ́ : Assign one violation for every stressed [ʌ] in an output.

c. FINALSTRESS: Assign one violation if the final syllable of a word is not stressed or if any non-final syllable of a word is stressed (abbreviated FINSTR).

d. ID-RD: Assign one violation for every output value of ROUND that differs from the corresponding input value.

e. ID-STR: Assign one violation for every output vowel where the level of stress from the corresponding input level.

2.5.1 Phonotactics in Extended Language

These constraints expand the possible typology of languages to include alternations in stress and vowel height, specifically opening up the possibility of vowel reduction with mobile stress in the paradigm. Importantly, these constraints also include a markedness constraint, *ʌ́, that must be dominated for the collective surface forms to be legal. For instance, in order for a surface form like *ʌ́k* 'oar' to be legal (map to itself, that is we temporarily assume the identity UR), the following ERCs must hold. Note that all of the constraints regulating obstruent voice from section 2.3 are still present, but have been supressed to avoid visual clutter.

(33)

ʌ́k	*ɔ	*ʌ́	FINSTR	MAX	DEP	ID-RD	ID-STR
a. ☞ ʌ́k		*					
b. ʌkɔ́		L			*W		*W
c. ɔ́k		L				*W	
d. ʌk		L	*W				*W

The full set of phonotactic ERCs for the language is the union of the ERCs established in section 2.3 and those immediately above. This set is displayed below:

	*ɔ	*ʌ́	FINSTR	*OBSVOI	*IVV	*OBSVOI#	MAX	DEP	ID-VOI	ID-RD	ID-STR
				L	L		W				
				L	L			W			
			L						W		W
			L							W	
			L	W							W

2.5.2 Morphophonology in the Extended Language

A particularly important ERC appears in the fourth line of the above table. This ERC shows that in order for the surface forms of the language to be legal, ID-ROUND must outrank *ʌ́. In effect, this rules out [ʌ-ɔ́] alternations being caused by avoidance of stressed [ʌ́]. More formally, this ERC is inconsistent with any morphophonological ERC set derived from a UR that contains [ʌ] in Figure 2.5. For example, consider the ERC set that is derived from attempting to ensure /ʌ́k/ maps to /ɔ́k/ 'mast':

(35)

ʌ́k	*ɔ	*ʌ́	FINSTR	MAX	DEP	ID-RD	ID-STR
a. ☞ ʌ́k						*	
b. ʌ́k		*W				L	
c. ʌkɔ́					*W	L	*W
d. ʌk			*W			L	*W

In the tableau above, the fully faithful loser [ʌ́k] can only be ruled out by the markedness constraint *ʌ́ outranking ID-ROUND, but it was just shown that in order for the observed word ʌ́k to be legal at all, precisely the opposite ranking must hold. Because the phonotactic ERC set

encodes all properties of the ranking that are necessary for surface forms as a whole to be legal, the narrower paradigmatic hypothesis cannot be true. In short, the faithful realization of all underlying round vowels is prioritized over the potential dispreference against their being stressed.

An attentive reader might object that the two above tableaux seek to map the same input to different outputs (the same input-output pair must be both a winner and a loser, a clear contradiction). The inconsistency that has been detected may therefore have more to do with OT languages being defined as functions than the impossibility of any particular featural change. This objection is ill-founded, because the ERCs merely reflect the content of the constraints which must be ranked. The phonotactic ERC set includes a statement that it is more important to protect an underlying [-ROUND] feature than it is to satisfy a surface constraint against stressed [ʌ́]. The morphophonological ERC set contains the exact opposite statement. It is this contradiction that underlies the extrapolation that low vowels cannot be the underlying source for vowels that alternate between mid and low heights, not the spurious fact that the same UR was being used to generate two different surface forms.

With the learner having encountered a contradiction in this branch of the search for constraints that can dominate ID-ROUND, this morphophonological hypothesis must be abandoned and a different UR must be considered. In this case, it works to make the allomorph found in ʌg-ʌ́ 'mast-acc' be unfaithful for the value of the feature ROUND. For instance, the UR /ɔg-ʌ́/ requires the following rankings to hold if ʌgʌ́ is its optimal surface form.

(36)

ɔg-ʌ́	*ɔ	*ʌ́	FINSTR	MAX	DEP	ID-RD	ID-STR
a. ☞ ʌgʌ́		*				*	
b. ɔgʌ́	*W	*				L	
c. ʌgɔ́		L				**W	
d. ɔ́gʌ		L	*W			L	**W
e. ɔ́g		L				L	*W

This set of ERCs is compatible with the phonotactic ERCs, which informally means that positing an underlying unstressed /ɔ/ is a viable extension of the phonotactic rankings.[18] Furthermore, as should be expected given the earlier discussion of morphophonological learning in the simplified language, the UR /ɔg/ can map to [ɔ́k] 'mast':

---

[18]The morphophonological ERCs in the tableau above are slightly incomplete, because contenders with unfaithful obstruent voice have been supressed. This is not a problem, since such candidates are losers for this allomorph, and the phonotactic ERCs do not rule out the rankings that make them lose.

(37)

ɔg	*C̥	*V́	FINSTR	*VOIOBS	*VOIOBS#	MAX	DEP	ID-LO	ID-STR	ID-VOI
a. ☞ ɔ́k									*	*
b. ɔg	*W		*W	*W	*W				L	L
c. ɔ́g				*W	*W				*	L
d. ɔk	*W		*W						L	*
e. ʌk			*W					*W	L	*
f. ɔ́						*W			*	L
g. ɔ́gʌ			*W	*W			*W		*	L

The phonotactic ERCs also impose an important restriction on the underlying voice specification of the morpheme 'mast'. The allomorph [ʌg-ʌ́] has a non-word-final obstruent, that is, an obstruent that is not in a potentially neutralizing environment given the phonotactic ERCs. The surface value is thus contrastive, ensuring that the underlying specification matches the surface value.

With all of the URs with voiceless obstruents or an underlying low vowel ruled out, the only URs remaining are /ɔg/ and /ɔ́g/. Under the constraint set used here, it does not matter which of these is used, as final stress can be generated from either UR. By considering all paradigmatic forms, the learner is able to build a set of grammars that allow all surface forms to be legal and does not rule out the alternations observed in paradigms. Under the available grammars, the only URs for some paradigms are composite URs.

38

2.6 Algorithm Pseudocode

To limit the amount of imprecision in the discussion of the basic process for finding phonological analyses from paradigmatically labeled data, I provide the following pseudocode summary of the intended method. The learning procedure can be productively conceived of as having two stages. The first stage, shown in algorithm 1, is concerned with collecting phonotactic and morphophonological ERCs. The second stage, shown in algorithm 2, is concerned with finding the rankings under which the allomorphs of all alternating morphemes are correctly generated. The pseudocode confronts several issues that were glossed over in the prose discussion. As a result, the algorithms are accompanied with a line by line walkthrough.

Algorithm 1 Collecting Constraint Rankings

1: **function** COLLECTION(C, T) $\triangleright C = \text{CON}, T = \text{text of paradigm labeled forms}$

2:     $Perc, Merc \leftarrow \emptyset, \emptyset$ $\triangleright$ Phonotactic, Morphophonological ERCs

3:     $A, M \leftarrow \emptyset, \emptyset$ $\triangleright$ Allomorphs, Morphemes

4:     **for** $f \in T$ **do** $\triangleright f = \langle str, \langle morphs \rangle, \langle allo\ indices \rangle \rangle$

5:        $Perc \leftarrow Perc \cup ercs(contenders(f[0], C), f[0])$

6:     **for** $f \in T$ **do**

7:        **for** $i \in range(length(f[1]))$ **do**

8:           $m \leftarrow f[1][i]$

9:           $M \leftarrow M \cup \{m\}$

10:          $obA \leftarrow f[0][f[2][i]]$ $\triangleright$ string representation of observed allomorph

11:          $A \leftarrow A \cup \{\langle m, obA, context(f[0], obA) \rangle\}$

12:          $Alts \leftarrow \{obA\}$

13:          **for** $a \in A$ **do** $\triangleright a = \langle morpheme, allomorph, context \rangle$

14:             **if** $a[0] = m$ **and** $a[1] \neq obA$ **then**

15:                $Alts \leftarrow Alts \cup \{\langle a[1], a[2] \rangle\}$

16:          $URs \leftarrow urs(disparities(Alts))$

17:          **for** $ur \in URs$ **do**

18:             $h \leftarrow \emptyset$

19:             **for** $alt \in Alts$ **do**

20:                **if** $ercs(contenders(ur + alt[1], C), alt[0] + alt[1])$ **then**

21:                   $h \leftarrow h \cup ercs(contenders(ur + alt[1], C), alt[0] + alt[1])$

22:                **else** $h \leftarrow \emptyset$

23:                   **break**

24:             **if** h **and** $consistent(h \cup Perc)$ **then** $Merc \leftarrow M \cup \{\langle m, h \rangle\}$

25:     **return** $\langle Perc, Merc, M \rangle$

2.6.1 Algorithm 1 Lines 1-3: Initialization

Lines 1-3 cover the starting values of the algorithm. The algorithm's arguments are a list of constraints C and a text T of paradigm-labeled surface forms. Lines 2 and 3 initialize the key variables for phonotactic ERCs $Perc$, morphophonological ERCs $Merc$, observed allomorphs A and observed morphemes M to empty sets.

2.6.2 Algorithm 1 Lines 4-5: Phonotactic Loop

In line 4, the algorithm starts to iterate over the paradigm labeled forms f held in the text T. As discussed in section 2.1.2, paradigm labeled data are a tuple consisting of a string, a tuple of morpheme identifiers, and a tuple of tuples containing the indices where each morpheme is realized in the string.

This loop populates $Perc$, the set of phonotactic ERCs. Following the discussion in section 2.3, line 5 augments the set of phonotactic ERCs. It does this by adopting the observed SR as an input and running the *contenders()* algorithm from Riggle (2004) on this input and the constraint set C. ERCs are extracted from the resulting set of possible winners with the function *erc(x, y)*, which takes a set x of candidates paired with violation vectors and a winner y from within x and produces vectors where absolute violation values are replaced with W, L or e. The goal is to ensure that the observed SR is not rendered illegal by the grammar winner, so the observed SR is designated as the winner.

It is important that $Perc$ be complete before morphophonological ERCs are collected. Otherwise a set of morphophonological ERCs can be judged to be consistent and stored, only for phonotactic ERCs collected from a later form to render that judgement incorrect. Hence, the for-loop terminates on line 5 and immediately restarts for the morphophonological loop.

2.6.3 Algorithm 1 Lines 6-24: Morphophonological Loop

At this point in the algorithm, the observed paradigmatically related form is broken down into its component allomorphs and different appropriations of unfaithfulness, as suggested by alternations

considered. Ultimately these appropriations of unfaithfulness are distilled into ERCs.

Line 6 starts this procedure by opening a for-loop over the morphemes observed in the form f. To facilitate reference to morphemes and the allomorph of that morpheme, the for-loop actually iterates over the position held by each morpheme in the tuple of morpheme identifiers in f.

Lines 8-12 manage variables pertaining to the current morpheme and allomorph. The morpheme that is currently being examined is assigned to the variable m, and since it has been observed, it is added to the set of observed morphemes M. Likewise, the current allomorph of the morpheme in question is assigned to obA, which, along with the morpheme identifier and the context in which the allomorph appears, is added to the set of observed allomorphs A. Finally, further processing will depend on the alternations that the morpheme undergoes, so a new variable $Alts$ is initialized as a set containing obA.

Lines 13-15 carry out the search for already observed alternants of the morpheme m. Every observed allomorph $a \in A$ is tested for whether it is an allomorph of m and if it is different from obA. If both tests are positive, the alternant is added to $Alts$.

As discussed in section 2.4, alternations between allomorphs of the same morpheme reveal what phonological features must be unfaithful to the input specification. However, there is no clue as to which surface value is faithful to the input. The learner must therefore try allocating different amounts of unfaithfulness to each observed allomorph. This is carried out in line 16. I assume a function *disparities()* identifies the locations of differences between allomorphs and what the differences are. The value this function takes on $Alts$ is passed to a function *urs()*, which constructs a set URs of underlying representations that reflect the various ways unfaithfulness could be allocated among surface allomorphs.

These URs are iterated over by the for-loop in lines 17-23. Each individual UR ur represents a different hypothesis of how the observed allomorphs might be unfaithfully derived, and thus might require a different ranking of constraints than any other UR. Accordingly, a new morphophonological hypothesis h is initialized as an empty set for each UR in line 18. By necessity, ERCs must be extracted from the mapping between ur and each observed alt, which is taken up in lines 20-21.

Importantly, it is possible for an allomorph to be harmonically bounded from a non-faithful

41

UR, in which case obtaining ERCs for the UR will be pointless. Thus, line 20 tests whether any ERCs are produced at all once ur is embedded in the context $alt[1]$ of the allomorph $alt[0]$ and $alt[0]$ in $alt[1]$ is specified as the intended winner.[19] If the intended winner is not harmonically bounded, then ERCs can be computed and stored in h (line 20).

However, if $alt[0]$ in $alt[1]$ is harmonically bounded from ur, then the value of *contenders()* will not contain it and *ercs()* will return a null value. In this case, there is no point in storing any ERCs associated with ur, so h is reset to the empty set (line 22) and the for-loop started on line 17 is subsequently exited.

The last analytic step of the algorithm is performed on line 24, where h is tested for whether it is non-empty and if $h \cup Perc$ is consistent. If so, the morphophonological hypotheses $Merc$ are updated to contain a tuple $\langle m, h \rangle$. The inclusion of the morpheme m for which the ranking information pertains will be relevant in algorithm 2, where it is necessary to ensure that the final rankings are compatible with some hypothesis for every morpheme. The algorithm concludes on the next line.

2.6.4 Algorithm 2: Identification

The second stage of learning consists of seeking the combinations of morphophonological ERCs that, together with the phonotactic ERCs, can generate the observed corpus. The pseudo-code that carries this out is presented in algorithm 2.

2.6.5 Algorithm 2 Lines 1-2: Initialization

Algorithm 2 takes as arguments the values returned by algorithm 1, namely the set of observed morphemes M, the set $Merc$ of tuples of morphemes and sets of ERCs, and the set $Perc$, which is the set of phonotactic ERCs. The only variable that needs to be initialized outside of the main for-loop is Fin, which will contain every combination of ERCs that are compatible with the text

---

[19]Lines 20-21 contain some abuses of notation and inaccuracies. First, $alt[0] + alt[1]$ is meant to convey the combination of an allomorph with its context, but the context ($alt[1]$) could easily be suffixal, prefixal or circumfixal. Second, and more importantly, $ur + alt[1]$ designates that the currently considered UR for the morpheme m is embedded in the surface context in which the allomorph $alt[0]$ appears. Strictly speaking, ur must be embedded not with $alt[0]$'s surface context, but with each of the URs for $alt[0]$'s surface context.

Algorithm 2 Identifiying Viable Constraint Rankings

1: **function** IDENTIFICATION($Perc, Merc, M$)
2:      $Fin \leftarrow \{Perc\}$
3:      **for** $m \in M$ **do**
4:          $kill \leftarrow$ True
5:          $temp \leftarrow \emptyset$
6:          **for** $hyp \in Merc$ **do**
7:              **if** $hyp[0] = m$ **then**
8:                  **for** $mainHyp \in Fin$ **do**
9:                      **if** $consistent(hyp[1] \cup mainHyp)$ **then**
10:                          $temp \leftarrow temp \cup \{hyp[1] \cup mainHyp\}$
11:                          $kill \leftarrow$ False
12:          **if** $kill =$ True **then return** $\emptyset$
13:          **else** $Fin \leftarrow temp$
14:      **return** Fin

T processed in algorithm 1. On line 2 Fin is initialized to be a set containing $Perc$.

2.6.6 Algorithm 2 Lines 3-13: Main Loop

Intuitively, the goal of this algorithm is to ensure that no morphophonological hypothesis (set of ERCs) is inconsistent with all other morphophonological hypotheses. To achieve this end, the algorithm loops through the set of observed morphemes M from the text T. Loosely speaking, if a set of ERCs associated with a morpheme is consistent with a set of ERCs associated with another morpheme, those ERC sets are unioned. As soon as no ERC set associated with a morpheme is consistent with any other collection of ERCs, the algorithm halts and announces that the entirety of the observe language cannot be generated with OT. We walk through this procedure in more detail below.

Each time the for-loop pulls out a new m in the for-loop on line 3, on line 4 the algorithm prepares for the possibility that none of the ERCs associated with m are compatible with the collective ranking requirements already encountered. The variable $temp$, which will hold the rankings that m is consistent with is subsequently initialized to the empty set.

2.6.6.1 Algorithm 2 Lines 6-11: Testing Morpheme Hypotheses

In line 6 the tuples hyp of a morpheme identifier n and a set of ERCs h from $Merc$ are looped over. Recall that h is a set of ERCs that can compel the alternations observed in the allomorphs of n. Line 7 ensures that only hypotheses that are associated with the current morpheme are examined.[20]

Upon reaching line 8, the algorithm loops over the ERC sets $mainHyp$ stored in Fin, and in line 9 checks them for consistency with the ERC set contained in $hyp[1]$. If this test is passed, $temp$ is updated to include $hyp[1] \cup mainHyp$, and $kill$ is switched off (since all of the morphemes considered thus far can be generated by a grammar).

2.6.6.2 Algorithm 2 Lines 12-13: Winding Down

Once the for-loop started on line 8 is exited, either no hypothesis hyp associated with m was consistent with any consistent combination of prior hypotheses, or some hyp was consistent. If the former, then the algorithm returns $\emptyset$ as its indication that no grammar will generate the patterns in T. If the latter, Fin is updated to $temp$ so that it contains the most recent set of consistent ERC sets. If the ERC sets in Fin are never incompatible with all sets of ERCs associated with some morpheme, the algorithm will terminate and return the set of viable ranking hypotheses.

2.7 Local Summary

This chapter has explored the algorithm that forms the core of a phonological learner in OT. The data a learner can reasonably expect to confront are surface forms that may be related morphologically. There are two main types of information that this data reveals about the grammar. First, morphological relations between forms can be ignored in phonotactic learning, where the learner seeks to find the constraints that dominate markedness constraints that disprefer observed surface forms. Second, if a morpheme has more than one allomorph, in morphophonological learning

---

[20]Note that this is the first (and fairly trivial) way that the language can be detected to be inconsistent. If the faithfulness violations suggested by the alternations of n are inconsistent by themselves or with $Perc$, n will not be represented in $Merc$ and thus no hypotheses will pass the test in line 7. The only opportunity to flip the value of $kill$ is on the more deeply embedded line 11, so any morpheme failing line 7 will trigger the announcement of inconsistency.

the learner seeks to find the constraints that dominate the faithfulness constraints that govern the alternating features.

The learner pursues the grammars that could generate the observed corpus by building a set of phonotactic ERCs and sets of morphophonological ERCs reflecting every underlying specification suggested by alternations. The sets of ERCs are then tested for consistency with each other, with the ultimate goal of compiling at least one consistent set of ERCs that allow all observed alternations and all observed surface forms to be generated.

This procedure has been demonstrated to work for small toy languages, and most importantly, demonstrates the learnability of languages that require composite URs. In chapter 3 we will show that an existing human language, Russian, that requires such URs is evidently diachronically stable. Chapter 4 will move the learning theory further to propose a response for when there is no consistent set of ERCs that allow the entire corpus to be generated.

CHAPTER 3

Russian: A Case Study of Composite Underlying Representations

Chapter 2 introduced the topic of learning a phonological grammar from only the information present in surface forms and their morphological relationships to each other. Particular emphasis was placed on the necessity of synthesizing information from multiple paradigmatically related forms, which raises the possibility of composite representations that bear features of distinct surface forms but are identical to none. The parade case (on which the toy language in chapter 2 was based) of a system that requires composite representations is Kenstowicz and Kisseberth's (1977, chapter 2) discussion of Russian. Kenstowicz and Kisseberth discuss oxytone (final stress) nouns like the word for 'pie', where *pirók* 'pie' is unsuffixed and receives stress on the stem, revealing the quality of the last stem vowel, though word-final devoicing obscures the underlying voicing of the stem-final obstruent. To obtain the voicing of the stem-final velar, the genitive singular form *pirǝg-á* 'pie-gen.sg' must be consulted, though the genitive singular cannot be relied on for all features of the UR, since the stem-final vowel has been reduced to [ǝ]. The underlying representation that best allows this paradigm to be generated is the composite UR /pirog/, with the /g/ coming from the genitive singular and the /o/ from the nominative singular.

While the analysis of the Russian word for 'pie' is straightforward, a variety of alternative hypotheses could be advanced to cover the small number of forms cited by Kenstowicz and Kisseberth (1977). For instance, a rule could map a stem final sequence [ǝg] in paradigms with alternating final stress directly to [ók]. Such an analysis would work well if most stems that end in /g/ have [ó] rather than [á] when the stem vowel is stressed. Convincing arguments that humans employ such traditionally dispreferred solutions have been made, see Albright and Hayes (2003) and further afield Hayes, Zuraw, Siptar and Londe (2009) have argued for a limited ability of humans

to adopt phonetically unnatural constraints. Indeed, the Single Surface Base hypothesis (Albright 2002; 2005; 2010, briefly reviewed in section 4.6) forbids composite URs and thus requires that such a non-traditional analysis be formed. Albright (2002:101-106) conducts a partial survey of the Russian data and concludes that the prospects for such an analysis being accurate may be quite good.

To ascertain whether such an analysis is plausible, this chapter surveys Zaliznjak (1977), an extensive analytic Russian lexicon commonly used in morphophonological investigations (cf. Linzen, Kasyanenko and Gouskova 2013). The primary focus is on whether the paradigms that appear to require composite URs share a small number of segmental traits, that would permit a non-composite UR analysis to capture the data. For instance, if all (or a very large proportion of) stems end in *əg* before a suffix and *ók* (and not *ák*) when unsuffixed, a composite UR analysis may not be the only suitable analysis for the data. While this survey is concerned with a very small corner of the lexicon, the results are fairly strongly in favor of a composite UR analysis.

Note briefly that what is not at issue is whether Russian generally enforces word-final devoicing or vowel reduction. Textbook treatments of Russian, like Kenstowicz and Kisseberth (1979:53-55) discuss how word-final devoicing is often transferred to the second language of Russian speakers, and that loanwords and neologisms uniformly obey word-final devoicing. Similarly, unstressed vowel reduction is virtually exceptionless in Russian, with the only potential counterexamples coming from loanwords and ecclesiastical terms, though Gouskova (2012:99) indicates that failure to reduce in these words is only a feature of highly conservative speakers. What is at issue is whether paradigms where all paradigmatic cells have been targeted by one of reduction or devoicing can be generated by positing rules that undo the neutralization.

An additional common question is whether a particular phonological phenomenon is under-attested in the lexicon. It appears that paradigms requiring composite URs are not under-attested in the lexicon relative to the paradigms attested in the language generally. As discussed in section 3.3, while paradigms with stress alternations are a minority in Russian (approximately 9% of the nouns in Zaliznjak 1977), and we are interested in the subset with a stem-final voiced obstruent and vowels that reduce when unstressed, the 142 stems that suggest a composite UR analysis fall in line with the lexical trends of Russian.

3.1 Word-Final Devoicing

To better understand the facts as they stand in modern Russian, I conducted a preliminary survey of Zaliznjak (1977), an analytical dictionary of the Russian lexicon. The citation forms in Zaliznjak's dictionary were converted to digital format and provided with full inflectional paradigms by Andrej Usachev.[1] Following the method used by Linzen, Kasyanenko and Gouskova (2013) on this lexicon, nouns were identified by the size of their paradigm, which is uniform across the lexicon since paradigm gaps were filled by Usachev. All analysis reported here was done with the aid of a Python script written by myself. The resulting list of 29,521 nominal paradigms was passed through a simple morphological analysis to identify the nominal stem. Five paradigms were discarded as their inflectional paradigms did not conform to the major declension types, bringing the number of analyzed nouns to 29,516. The table below gives counts for the number of nominal stems according to whether they end in a vowel, sonorant consonant or obstruent.

(38)　　　Stem-Final Segments

Vowel	Sonorant	Obstruent
9,679	5,656	14,181

The class of stems of current interest are those that end in obstruents, and among those, the ones that occur in declension classes where the stem-final obstruent is placed in word-final and word-medial position, as in *kabantʃʲikʲ* 'pylon', *kabantʃʲikʲ-i* 'pylon-nom.pl'. Noun stems overwhelmingly occur in paradigms where the morphology moves the final consonant in and out of of the devoicing environment, with 14,012 of 14,181 stems falling into this category. The key comparison within this set is between the stems that end in voiced obstruents and those that end in voiceless obstruents. (39) summarizes the respective counts; because affricates and [x] are not opposed by voiced segments, (39) also displays the count for voiceless obstruents that have a voiced counterpart.

---

[1] The source file for the dictionary can be found at `www.speakrus.ru/dict/all_forms.rar`.

(39)   Stem-Final Obstruents

Voiced	All Voiceless	Paired Voiceless
3,308	10,704	9,516

Non-alternation is clearly the majority pattern in the lexicon as a whole. Figure 3.1 breaks down the consonant alternations by voicing pair, showing in each case whether the voiceless (unalternating) phoneme is more common than the voiced (alternating) phoneme. The conclusion that emerges from inspecting Figure 3.1 is that the bulk of the non-alternating segments are [k], while labial and alveolar places of articulation show stronger tendencies for voicing alternations.

3.2 Vowel Reduction

Russian vowel quality is correlated with consonant palatalization/velarization and stress (Avanesov 1956; 1972, Crosswhite 1999). In stressed syllables, five vowels are distinguished as shown in Figure 3.2, with the quality of the high front vowel varying between [i] and [ɨ] after palatalized and velarized consonants. Following Padgett (2001), non-back high vowels will be transcribed as [i]. When necessary to distinguish between high non-back vowel quality or when palatalization/velarization departs from the typical association with front/back vowels palatalization/velarization will be marked on the preceding consonant.[2]

Padgett and Tabain (2005) provide the examples of the Russian stressed vowel inventory in (40), where an evident transcription error in the word for 'law' is corrected:

(40)   'dˠim   'smoke'   'vʲit   'species'

        'sudnə   'ship'   klʲutʃ͡   'key'

        't͡sˠex   'workshop'   'delə   'business'

        'ɡot   'year'   'slʲos   'tears (gen.pl)'

        'pravə   'law'   'rʲat   'row'

---

[2]The form t͡sˠex 'workshop' in (40) does not have palatalization of the initial consonant, though it appears to be a loan word. The native Russian vocabulary palatalizes consonants before [e], though not necessarily across prefix boundaries.
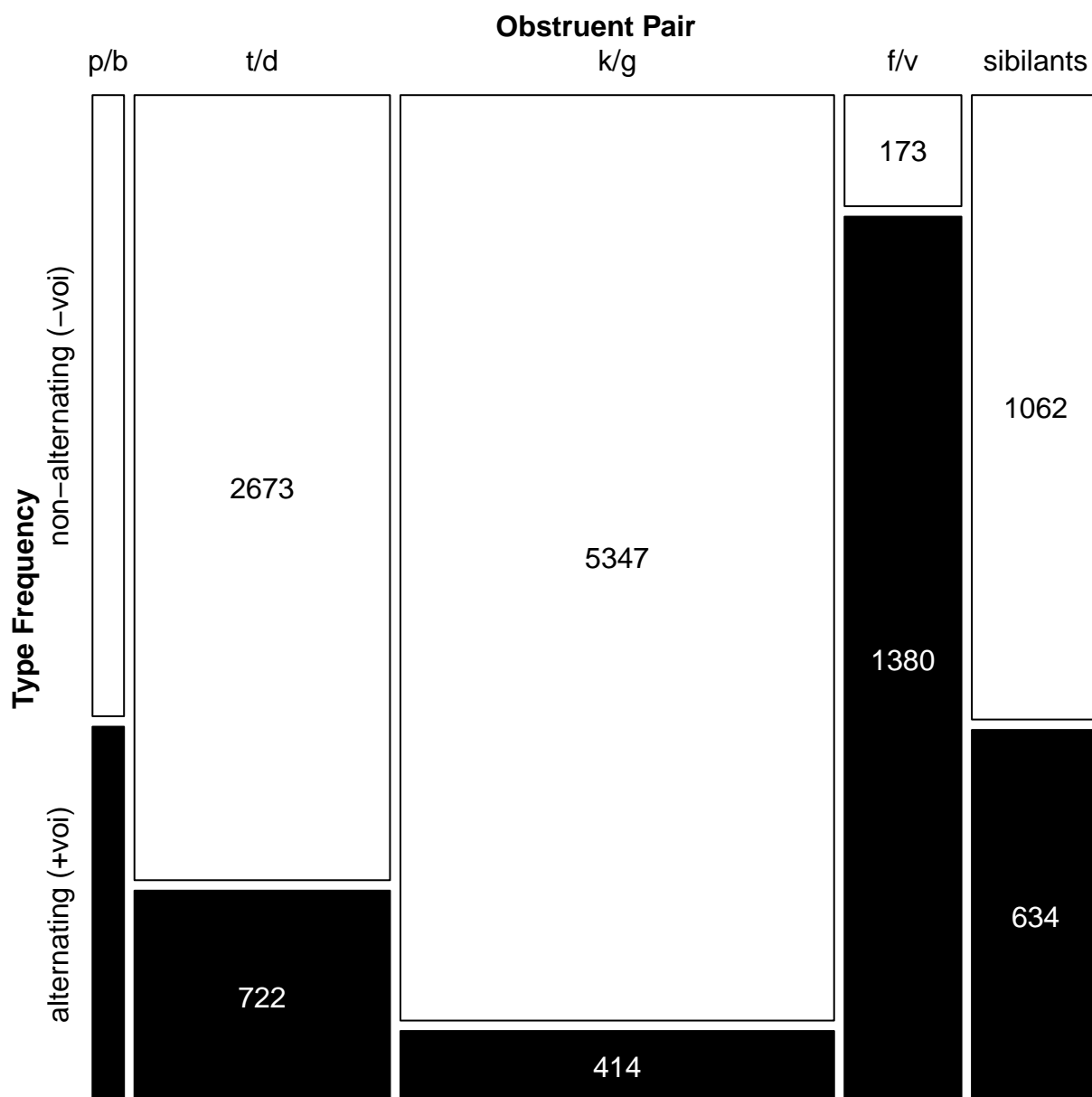
Figure 3.1: Prevalence of alternation versus non-alternation by voicing pair in nouns from Zalizn-jak (1977). Alternation is better attested relative to non-alternation at labial and alveolar places of articulation and in fricatives. The counts from Zaliznjak (1977) for [p, b] are 261 and 158, respectively. The sibilant pairs [s, z] and [ʃ, ʒ] have very similar rates of alternation versus non-al-ternation, and so were collapsed.

```
      (i) ——— (ɨ) ——— u

           e            o

              a
```
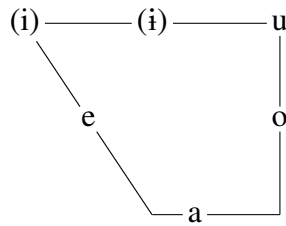
Figure 3.2: Russian Stressed Vowel Inventory

In unstressed syllables, the vowel inventory is reduced in different ways depending on the palatalization of the onset and position relative to stress. When the preceding consonant is palatal-ized, only two vowels, [i, u] are legal. Paradigmatic alternations show that the non-high vowels [e, o, a] all neutralize with [i] in this environment, as can be seen in the adjectival forms of the palatalized forms from (40) in (41), also drawn from Padgett and Tabain (2005).[3]

(41)   vʲid-ɐˈvoj   'species-adj'

       klʲutʃ-iˈvoj   'key-adj'

       dʲil-ɐˈvoj   'business-adj'

       slʲizɐtɐˈtʃivˠij   'tear gas (adj)'

       rʲid-ɐˈvoj   'row-adj (rank and file)'

The vowel inventory in unstressed syllables with velarized onsets shrinks to three vowels, [i, u] and a low vowel, from the full five vowel inventory. The phonetic character of the low vowel varies according to the position relative to stress. Immediately pre-tonic low vowels are higher than stressed low vowels, being described as [ɐ], while unstressed low vowels elsewhere are raised further to [ə]. The vowel system in velarized contexts is shown in Figure 3.3.

With three vowel categories after velarized consonants, the pattern of neutralization shifts. Where in palatalized contexts all non-high vowels are neutralized with [i], in velarized contexts only [e] is neutralized with [i]. Hence, Padgett and Tabain (2005) provide the examples reproduced in (42), which can be compared to the parallel non-adjectival forms in (40).

---

[3]The examples in (41) show that derivational morphology can trigger stress shifts and allow voiced consonants to surface faithfully. The derivational patterns that a noun stem takes part in could also therefore provide data that would motivate a composite UR analysis. Derivational morphology will not be considered here, primarily because of inconveniences posed in extracting such data from Zaliznjak (1977). Note, however, that this does allow the often vexed issue of the productivity of derivational morphology to be sidestepped.
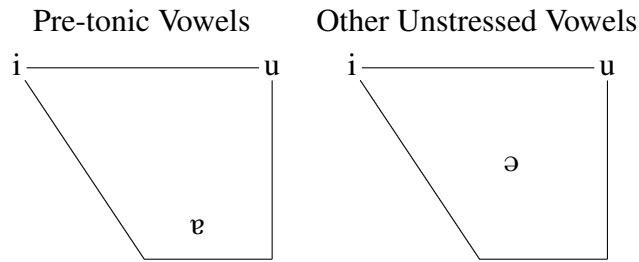
Figure 3.3: Russian Unstressed Post-Velarized Vowel Inventory

(42)   dʲim-ɐˈvoj   'smoke-adj'

        sud-ɐˈvoj   'ship-adj'

        t͡sʲix-ɐˈvoj   'factory shop-adj'

        gəd-ɐˈvoj   'year-adj (yearly)'

        prəv-ɐˈvoj   'law-adj (legal)'

Henceforth, the distinction between pre-tonic and otherwise unstressed vowels will not be relevant to our purposes. Transcriptions will mark reduced back vowels in velarized contexts with [ə]. The crucial point is that there are four neutralizing vowel reduction alternations in Russian: /o/ neutralizes with /a/ in velarized contexts, /o/ and /a/ neutralize with /i/ in palatal contexts, and /e/ neutralizes with /i/ in both velarized and palatal contexts.

3.2.1   Stress Alternations

Our interest in vowel reduction is rooted in the alternations it causes when stress shifts between syllables in an inflectional paradigm. As mentioned above, the vast majority of nouns in Russian have non-alternating (columnar) stress, and only about 9% of all nouns in Zaliznjak (1977) have mobile stress. Stress alternations come in two types. The most common type is made up of nouns that exclusively have stress on the last syllable of the word. The other stress shifting nouns, accounting for approximately 1% of nouns in Zaliznjak (1977) alternate stress between the first syllable of the word and the last syllable of the word. An example of each type is provided in (43).

(43) UR nom.sg gen.pl gloss

/bagrets͡ˠ/ bəgréts͡ˠ bəgrʲits͡ˠ-óf 'scarlet'

/golovˠ/ goləvˠ-á gəlófˠ 'head'

Of the mobile stress words, approximately 640 have a stressed vowel that deletes in other members of the paradigm, as in *són* 'sleep', which has the genitive *sn-á* 'sleep-gen.sg'. While it is true that stress is mobile, if the stress then falls on an affix, the paradigm does not fit our specifications, as our interest is in stress conditioning vowel quality alternations. Removing these nouns leaves 2,038 nouns with relevant stress alternations. These nouns were then analyzed for which vowels had both stressed and unstressed allophones in the paradigm. Counts were tabulated both for the total number of vowels belonging to a particular phoneme and for the total number of stems containing vowels of the phoneme in question. These counts are presented in (44). Note that summing each row does not equal the total number of stems in the category, since the existence of alternations between the first and last syllable of the stem means that a single stem can contribute more than one vowel to the count, or even be categorized in more than one column. Non-parenthesized counts represent totals for a phoneme without accounting for the palatalization/velarization of the preceding consonant, parenthesized counts represent the total for the phoneme in a particular consonantal context.[4]

(44) Stressed Vowel Counts

	a	(Cʲa)	o	(Cʲo)	e	i	(Cˠi)	u
mobile vowels	611	(154)	398	(97)	277	540	(92)	333
mobile stems	604	(154)	388	(97)	273	538	(91)	333

Our chief concern is paradigms that suggest composite URs, which include alternating vowels withing mobile stress paradigms. Figure 3.2.1 illustrates how well attested alternating vowels are in mobile stress paradigms relative to the non-alternating categories they are neutralized with. Recall

---

[4]Our ability to detect consonantal environment is slightly hampered by our reliance on orthographic data. Russian orthography indicates palatalization of consonants with specialized vowel graphemes for back vowels and [i], but has only one grapheme for [e]. This is not generally a problem, since [e] occurs almost exclusively after palatalized consonants, but as (40) shows, there are some cases where it occurs after velarized consonants. Avanesov (1985:663) includes some words where the orthography indicates velarized consonants, but the spoken language has palatalized consonants.

that [o] and [a] neutralize with [i] in palatalizing contexts and [o] is mapped to [a] in velarized contexts. Figure 3.2.1 represents this by comparing the [a]-[o] alternation only against the [a]'s that occur in velarized contexts, while the [a]-[i] alternation and the [o]-[i] alternation is compared only against the post-palatal allophone of [i]. No such distinction is made for [e], which alternates with [i] in all contexts. Due to the tight link between front vowels and palatalization in Russian, we can assume that the bulk of the alternations are in the palatal context, though as mentioned above, the orthography does not allow us to be certain.

From Figure 3.2.1 we can conclude that the [a]-[o] alternation and the [i]-[e] alternation are the most well attested relative to their non-alternating counterparts. This is not surprising, as the trend in favor of [a]-[o] and [i]-[e] alternations reflects the historical link between front vowels with palatalization and back vowels with velarization.

3.2.2 Low [e, o] Counts

A brief look at (44) makes it immediately clear that in mobile stress paradigms, [o] and [e] (the vowels that undergo neutralizations in every context when unstressed), are the third and fifth most numerous vowels, while [a] and [i] (the vowels with which [o] and [e] are neutralized) are the first and second most well-attested vowels. A tempting explanation for this is that the language has been systematically, if slowly, losing alternating mid vowels in favor of the non-alternating phonemes that they are neutralized with. While it is certainly plausible that learners exposed to any neutralizing alternation could adopt the neutralized value as the underlying one, a broader look at the language indicates that this is unlikely to be the main cause of the low counts for [e] and [o].

If it is the neutralizing allophony that is driving the low numbers of [o] and [e] (and high numbers of [a] and [i]) in mobile stress paradigms, then columnar stress paradigms, where such allophony is not present, are a suitable baseline. (45) presents the counts for mobile stress paradigms alongside parallel data for columnar stress paradigms. The columnar stress figures include mobile stress paradigms where the stressed stem vowel is deleted.[5]

---

[5]Furthermore, the number of columnar stems is two greater than would be expected, since the suppletive paradigms for *t͡ʃelovek* 'man' and *rebʲonok* 'child' were counted as both columnar and mobile. The singular forms of these words have columnar stress, while the plural forms have mobile stress, viz. *lʲúdʲ-i* 'man-nom.pl' and *lʲudʲ-mʲí* 'man-instr.pl'. Figures for [a] in columnar stress paradigms include palatalized and velarized contexts, but the count for [a]
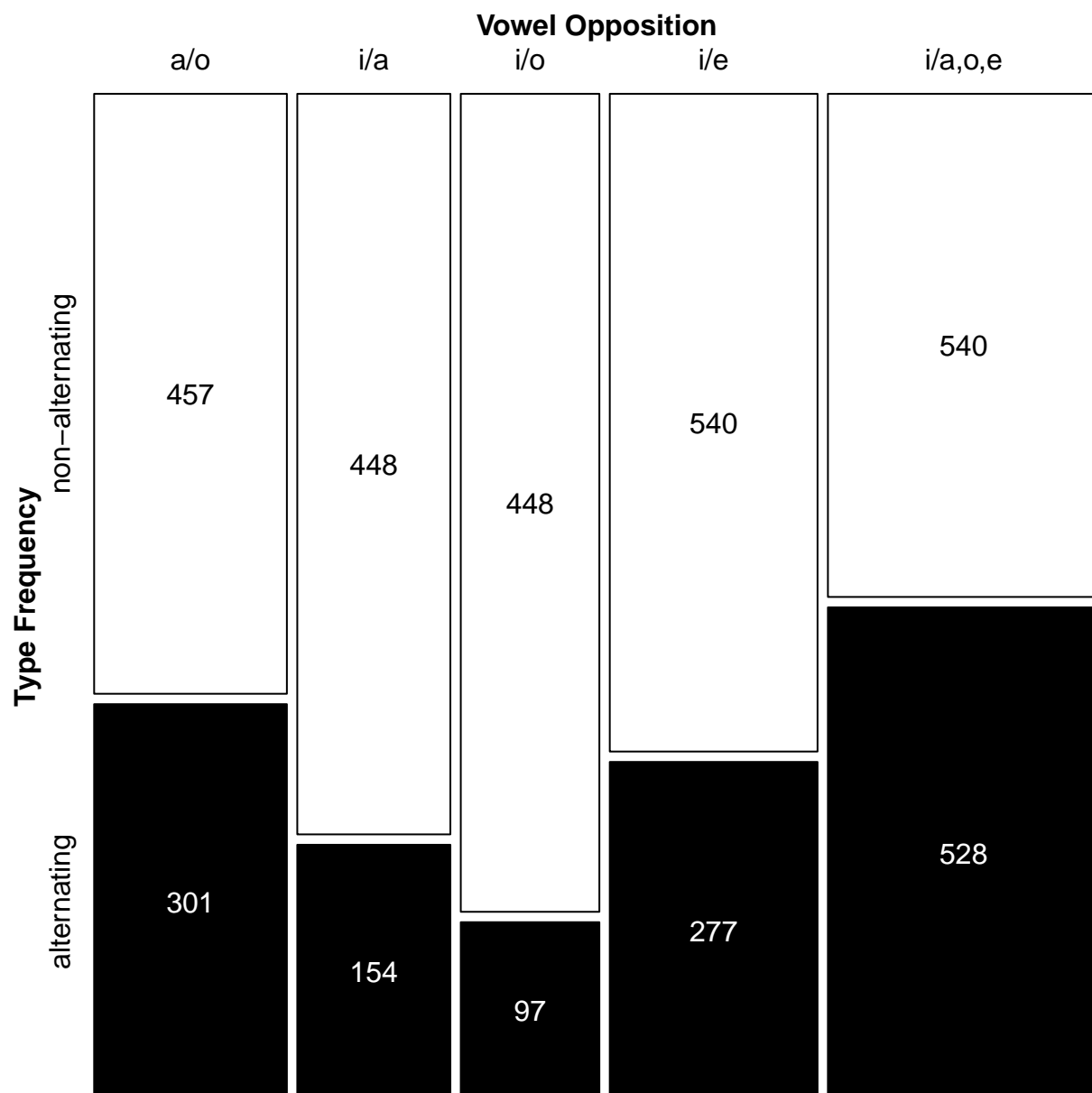
Figure 3.4: Prevalence of alternation versus non-alternation by vowel pair and preceding conso-nantal context in nouns from Zaliznjak (1977). The value for [i] in the /a/ versus /i/ and /o/ versus /i/ comparison represents instances of /i/ in post-palatal environment, while the /e/ versus /i/ com-parison includes all contextual variants of [i]. Alternation is best attested between [a] and [o], followed by [i] and [e]. If all contextual variants of [i] are pooled, then nearly half of all types of unstressed [i] are mapped to one of [e], [o] or [a].

(45)				Stressed Vowel Counts					
	a	(Cʲa)	o	(Cʲo)	e	i	(Cʸi)	u
columnar vowels	7,861	—	7,309	(900)	5,914	5,751	(585)	1,752
columnar stems	7,705	—	7,221	(900)	5,825	6,266	(578)	1,746
mobile vowels	611	(154)	398	(97)	277	540	(92)	333
mobile stems	604	(154)	388	(97)	273	538	(91)	333

Figure 3.5 plots the non-parenthesized totals of the columnar and mobile vowel counts in (44) as percentages of the respective paradigm types. The figure confirms the general impression: [o] and [e] are a greater share of the stressed vowels in columnar stress paradigms than in mobile stress paradigms, and in mobile stress paradigms [a] and [i] are a greater share of stressed vowels than in columnar stress paradigms.

However, the 7.2 percentage point drop in the prevalence of [o] and the 7.9 percentage point drop in the prevalence of [e] in mobile paradigms are not matched by equivalent gains in [a] and [i], which rise by 0.8 percentage points and 4.9 percentage points, respectively. Rather, the largest jump in attestation comes from [u], whose rate of attestation as a stressed vowel in mobile stress paradigms is 9.3 percentage points higher than in columnar stress paradigms. Importantly, [u] does not share unstressed allophones with any of the vowels in question. As a result, the skewing of the vowel counts in mobile stress paradigms relative to columnar stress paradigms cannot be due entirely to alternations weakening mid vowels at the expense of the phonemes they are neutralized with. It is more likely that the counts of vowels in the relatively small number of mobile stress paradigms were skewed to begin with, in which case these data will not permit us to assess the effect of alternations weakening the counts of mid vowels. An alternative possibility is that some other process has asymmetrically targeted mid vowels in mobile stress paradigms (perhaps transferring their stems to columnar stress paradigms).

_____

in palatalized contexts was not tabulated independently.
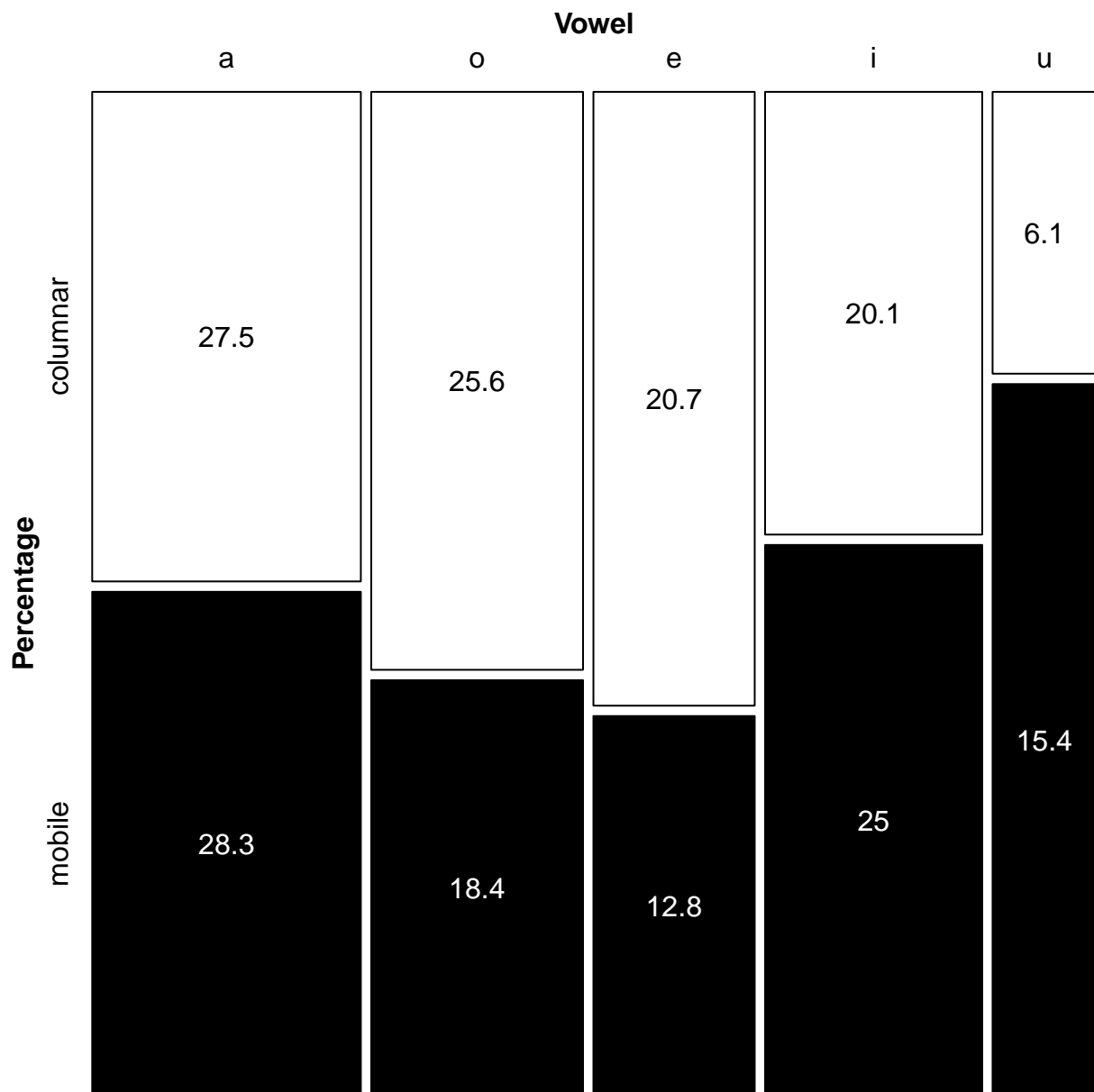
Figure 3.5: Individual vowels as percentages of the total number of stressed vowels in mobile and columnar stress paradigms. In mobile paradigms [e] and [o] are less robustly attested than in columnar stress paradigms, while [a], [i] and [u] are more robustly attested in mobile stress paradigms than in columnar stress paradigms. The gains for [i] and [a] in mobile stress paradigms cannot have been solely due to [o] and [e] losing lexemes to [i] and [a], as [u] is the category with the bulk of the gain relative to columnar stress.

3.3 Devoicing and Reduction

Having reviewed the general phenomena of vowel reduction and word-final devoicing, we turn now to the paradigms where the alternations caused by these processes suggest a composite UR analysis. As should be expected from the small numbers of paradigms that have stress alternations, paradigms that end in voiced consonants and have vowels that alternate are not particularly numerous, with 142 mobile stress paradigms meeting the criteria. Importantly, however, the counts from sections 3.1 and 3.2 allow us to estimate the probability that a Russian noun stem will end in a voiced obstruent and the probability that a mobile stress paradigm will contain an alternating vowel. The product of these probabilities times 2,038 (the number of mobile stress paradigms) is how many paradigms that suggest a composite UR we might expect to observe. As (46) shows, the number of observed composite paradigms is generally not lower than expected with these probabilities.

(46)   Observed vs (Expected) Composite Paradigms

		b		d		g		v		z		ʒ	
		0.5%		2.2%		1.3%		4.2%		1.4%		0.6%	
Cˠo	13.9%	7	(1.4)	20	(6.2)	10	(3.6)	27	(11.9)	5	(3.9)	6	(1.6)
e	12.8%	2	(1.3)	8	(5.7)	4	(3.3)	12	(10.9)	0	(3.6)	5	(1.4)
Cʲa	7.1%	1	(0.7)	1	(3.2)	0	(1.8)	0	(6.1)	2	(2)	5	(0.8)
Cʲo	4.5%	1	(0.4)	2	(2)	0	(1.2)	3	(3.8)	1	(1.3)	20	(0.5)

(46) demonstrates that it is unlikely that a dispreference against composite URs has whittled away the number of composite paradigms over time. If anything, the number of composite paradigms has increased relative to our baseline expectation, though the robustness of composite paradigms is skewed towards paradigms like *vʲiʃéstʃ* 'stuff (gen.pl)', *vʲiʃʲistv-ó* 'stuff-nom.sg', or *pəlubók* 'demigod (nom.sg)', *pəlubəg-á* 'demigod-gen.sg', which contain [o] or [e].

The more pressing issue is the viability of a non-composite UR analysis of these paradigms.[6] For concreteness, I will assume that a non-composite UR analysis follows Albright's (2002) Single

---

[6]To be sure, 142 paradigms are probably not beyond the capacity of a person to memorize, so in some sense, a non-composite UR analysis that relies on lexical listing of morphological alternants must be viable.

Surface Base hypothesis (reviewed in section 4.6) by picking some morphophonological category (e.g. the genitive plural, a suffixed form or an unsuffixed form), and using it as the input to all phonological rules. The key to success for such a model is that the neutralized category in the allomorph chosen must be the only category in that phonological environment. It is no good to form a process turning [ə] into [o], if half of the [ə]'s in that phonological environment must be mapped to [a].

Figure 3.6 illustrates the type frequencies of composite paradigms with an [o]-[a] alternation against non-composite paradigms with non-alternating [a], categorized by which voiced obstruents the stems end in. It depicts the answer to the question "how accurate would processes that restored [o] from [ə] according to the stem-final consonant be?". As can be seen in figure 3.6, the answer is "fairly accurate", since non-alternating [a] paradigms predominantly have different stem-final consonants than paradigms with an [o]-[a] alternation. Concretely, a process that took a suffixed form as an input and mapped [əd] to unsuffixed [ót] would get the right answer 20 out of 26 times, and similar accuracy figures exist for the other voiced obstruents. The mistakes such a process would make would be in over-extending [o] in paradigms where [a] is correct. (47) exemplifies an application of this process to forms where it would derive the correct and incorrect results, respectively.

(47)

	'movement'	'mood '	Gloss
	xəd-á	ləd^ɣ-i	UR Source (nom.pl)
	/xəd/	/ləd^ɣ/	UR
	xót	lót^ɣ	əd → ót / _ #
	[xót]	[lót^ɣ]	SR
	[xót]	[lát^ɣ]	Correct SR

However, selecting a suffixed form and undoing the vowel neutralization is much less straightforward if the vowels in question are neutralized to [i]. This is primarily because, as Figure 3.7 shows, the non-alternating [i] is roughly as well attested throughout the space of voiced obstruents as the main alternating category [e]. Further, there is the problem of having to correctly specify the stressed vowel, since it is in fact a four-way neutralization between /e/, /i/, /a/, /o/. In fact, for every
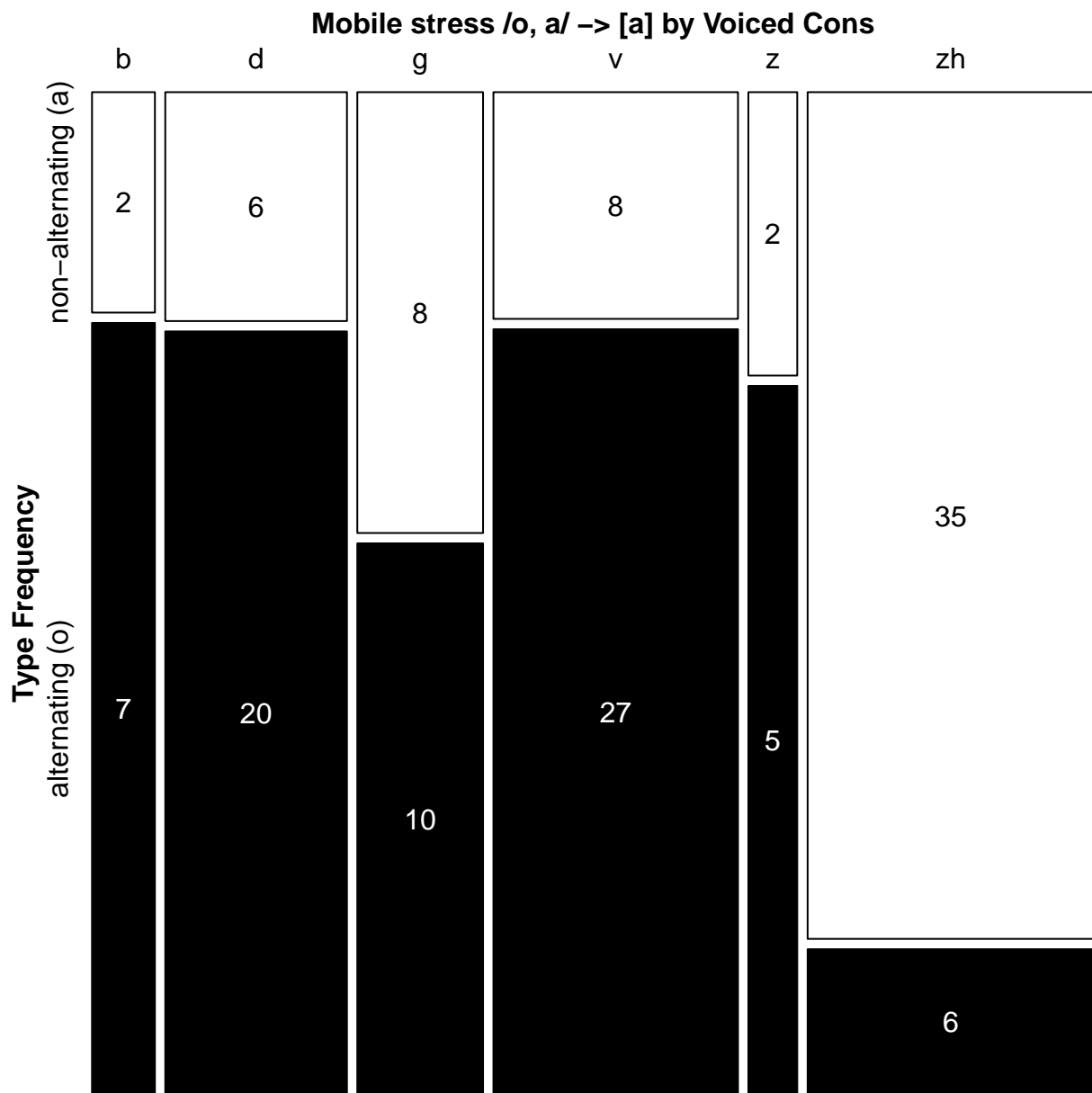
Figure 3.6: Prevalence of unstressed [ə] that alternates with stressed [ó] against unstressed [ə] that alternates with stressed [á] by stem-final voiced consonant in nouns from Zaliznjak (1977). Alternation with [ó] is generally better attested than alternation with [á], except for when the stem-final consonant [ʒ].

voiced obstruent except /z/, restoring the most well attested stressed alternant from unstressed [i] will be correct at most only slightly over 50% of the time. (48) provides a schematic example of the pitfalls.

(48)

	'hubbub'	'case	'waxwork'	'swift (bird)'	Gloss
	gəldʲiʒʲ-í	pədʲiʒʲ-í	mulʲiʒʲ-í	strʲiʒˠ-í	UR Source (nom.pl)
	/gəldʲiʒʲ/	/pədʲiʒʲ/	/mulʲiʒʲ/	/strʲiʒˠ/	UR
	gəldʲóʃʲ	pədʲóʃʲ	mulʲóʃʲ	strʲoʃ	Cʲiʒ → Cʲóʃ / _ #
	[gəldʲóʃʲ]	[pədʲóʃʲ]	[mulʲóʃʲ]	[strʲoʃ]	SR
	[gəldʲoʃʲ]	[pədʲéʃʲ]	[mulʲáʃʲ]	[strʲiʃ]	Correct SR

Turning the tables so that the process attempts to restore voicing to devoiced consonants runs into similar problems. For instance, as shown in Figures 3.8 and 3.9, there are a greater or equal number of stem-final non-alternating voiceless consonants than alternating voiced consonants in mobile stress paradigms with a stressed [o] or a stressed [e]. Any process with a chance at being accurate would have to impose a non-alternating voiceless obstruent over the original voiced obstruent. The effect of this would be to remove composite paradigms from the language.

One set of composite paradigms where it is fairly feasible to restore voicing from an allomorph where it has been devoiced are those where the stressed vowel is [o] in a palatalized context. This is shown in Figure 3.10. Composite paradigms where the stressed vowel is [a] in a palatalized context are too few to make a coherent graph, and so are not displayed.

By now it should be clear that attempting to generate the composite paradigms of Russian without a composite UR would require developing highly idiosyncratic processes. Such processes are not out of the question, but without reinforcement from high token frequency or a background of phonological or semantic reliability, it is fairly likely that paradigms with what would essentially be irregular phonology would be regularized at some point in the diachronic development of the language. Strikingly, following Kiparsky (1979) and Lunt (1980), the reduction and devoicing alternations in Russian have coexisted for approximately 700 years. If learners were treating the composite paradigms as anomalous, one might expect regularization to have already taken place. By contrast, if composite paradigms pose no special challenge to learners, the stability of these
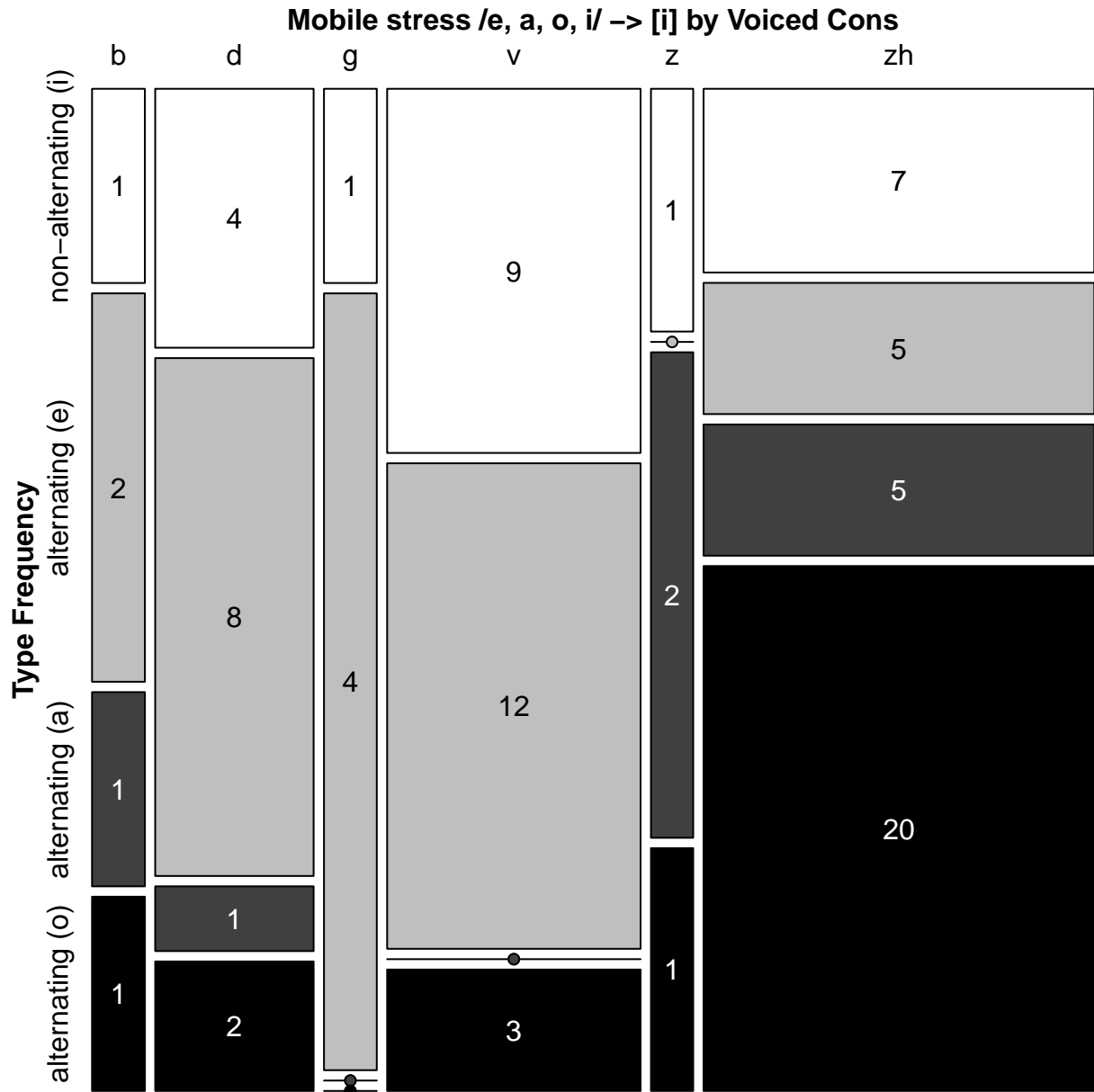
Figure 3.7: Prevalence of unstressed [i] that alternates with stressed [í, é, á, ó] by stem-final voiced consonant in nouns from Zaliznjak (1977). Alternation with any single vowel quality is typically offset by alternation with the remaining three vowel categories.
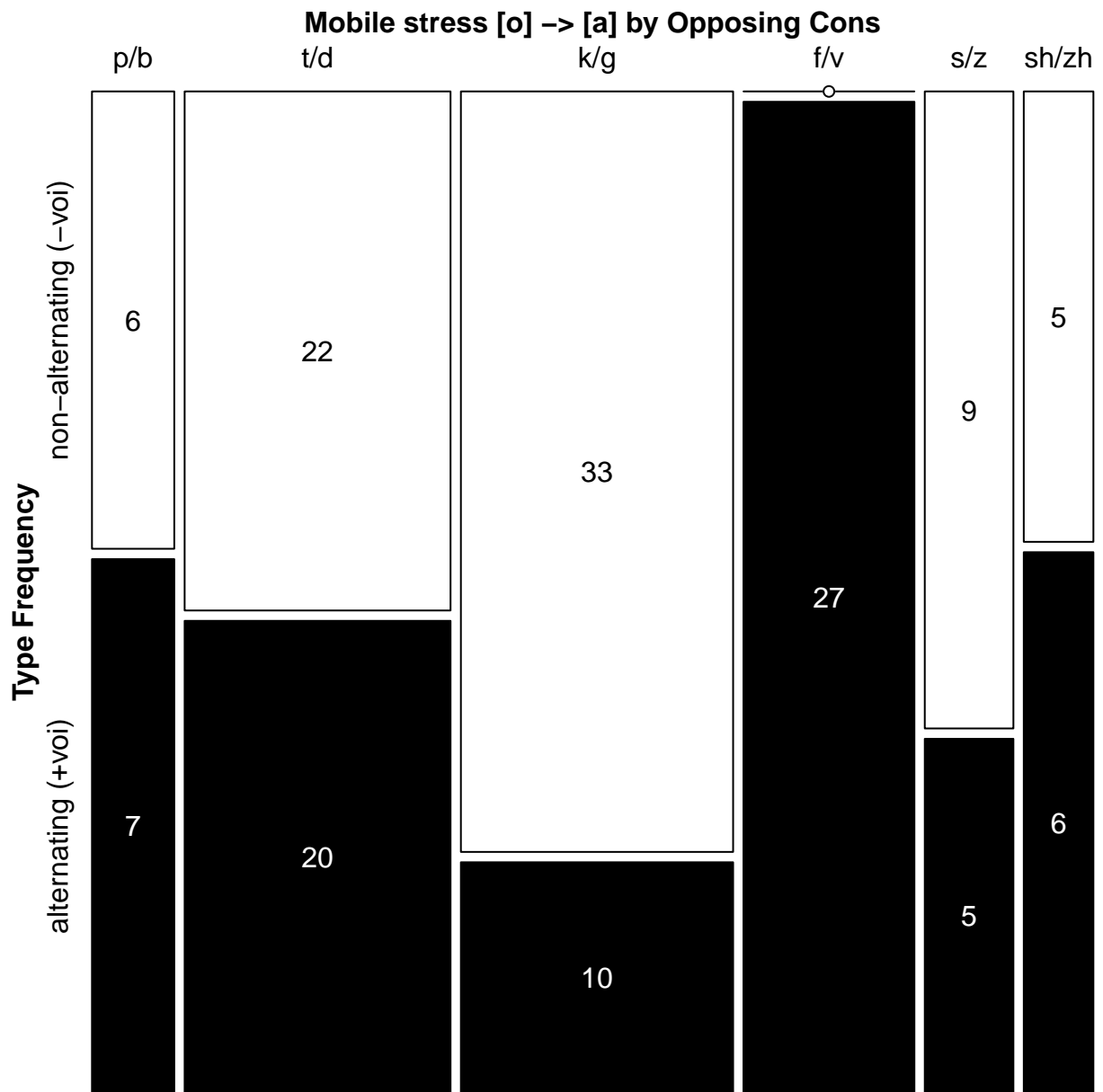
Figure 3.8: Prevalence of voiceless obstruents in word-final position that have voiced or voiceless realizations by obstruent type in mobile stress nouns with a stressed [ó] from Zaliznjak (1977). Alternating (voiced) and non-alternating (voiceless) obstruents are approximately evenly attested.
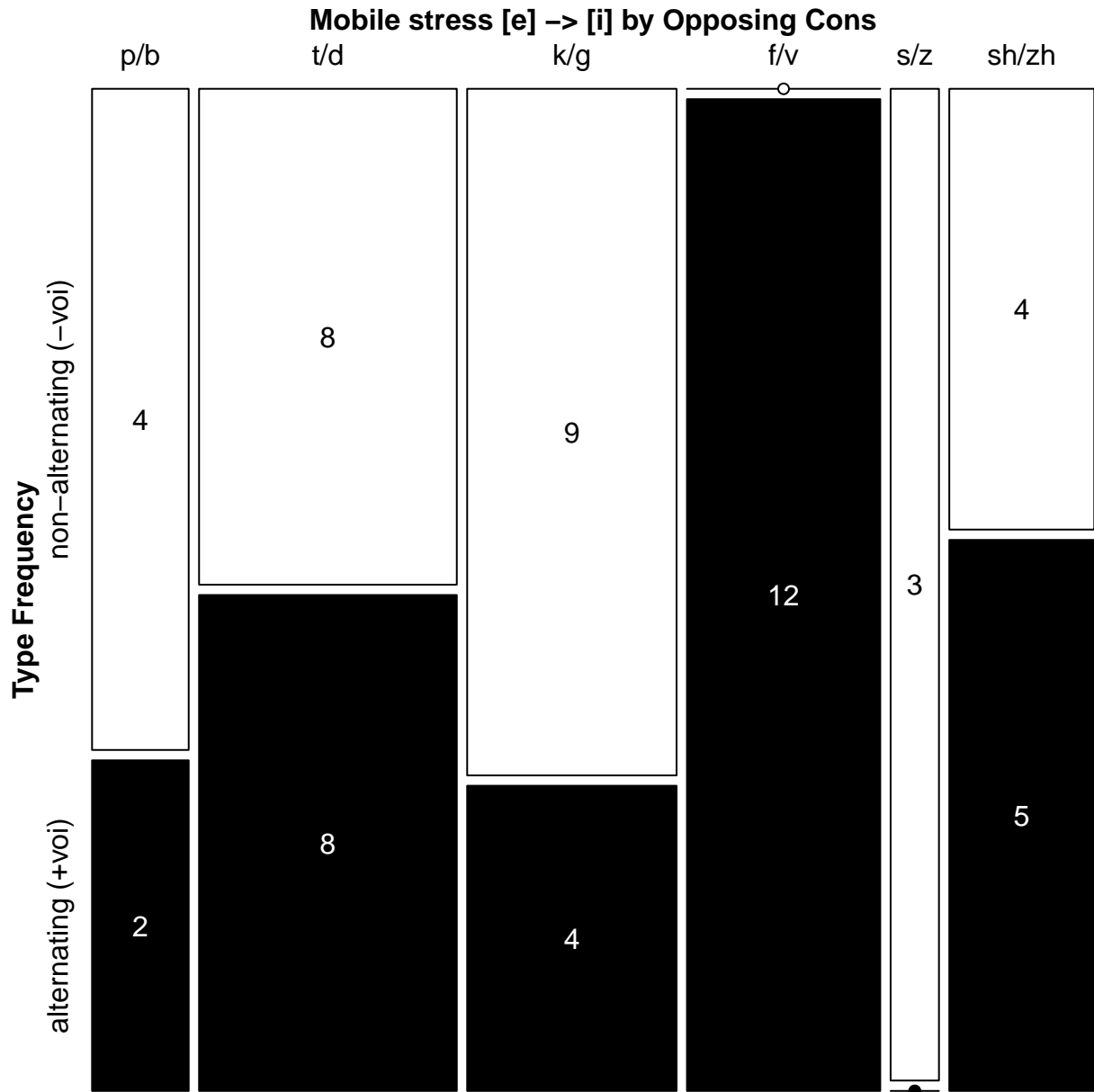
Figure 3.9: Prevalence of voiceless obstruents in word-final position that have voiced or voiceless realizations by obstruent type in mobile stress nouns with a stressed [é] from Zaliznjak (1977). Non-alternating (voiceless) stops are better attested in than alternating (voiced) stops.

Mobile stress [o] –> [i] by Opposing Cons
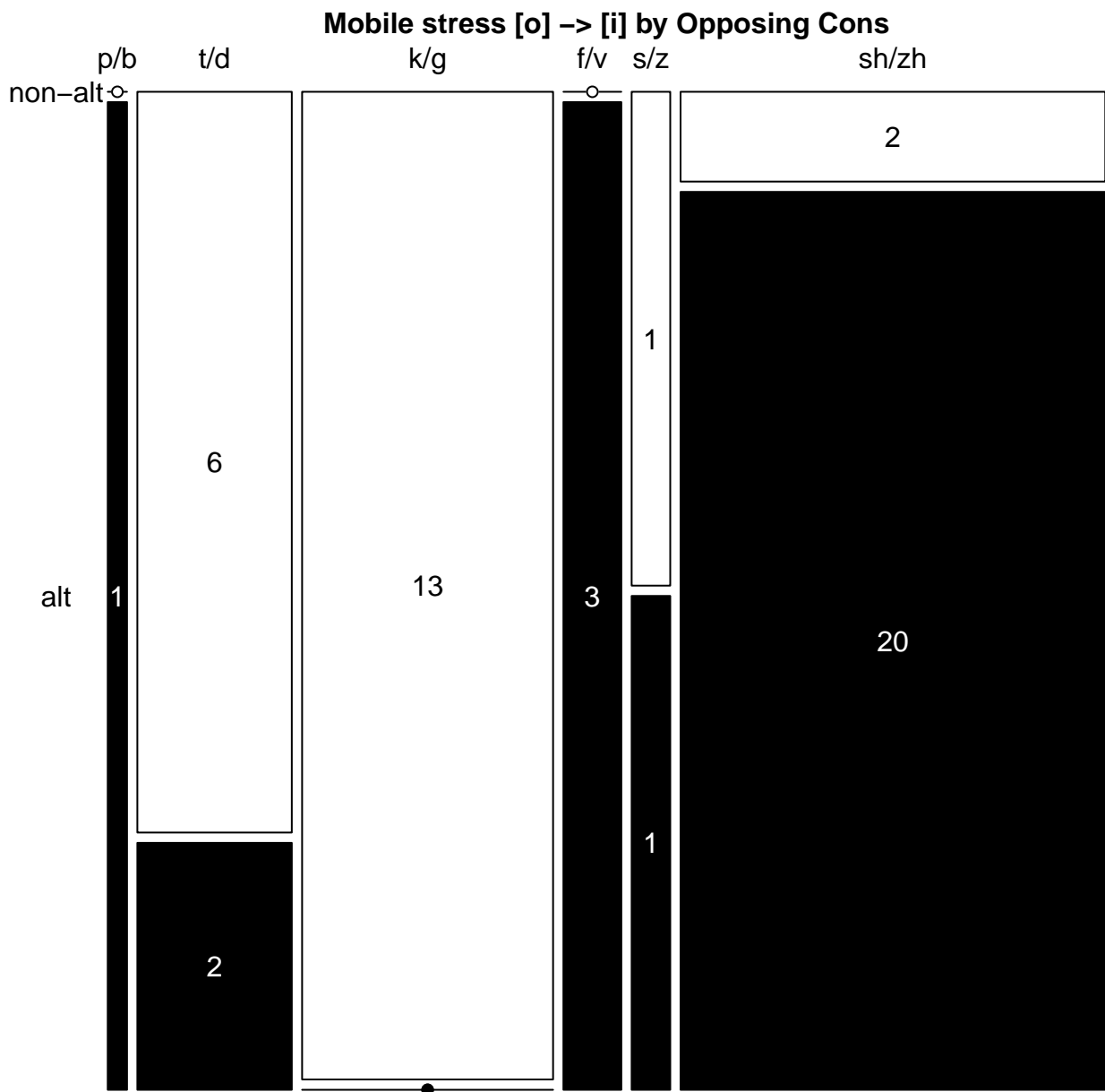
Figure 3.10: Prevalence of voiceless obstruents in word-final position that have voiced or voiceless realizations by obstruent type in mobile stress nouns with a stressed [ó] following a palatalized consonant from Zaliznjak (1977). Fricatives are heavily skewed towards being alternating (voiced), while stops are skewed towards being non-alternating (voiceless).

paradigms is unremarkable. As a result, a learning algorithm that can acquire composite URs, as discussed in chapter 2 is supportable.

CHAPTER 4

Imperfect Learning

4.1 Learning and Language Change

Language change is at first blush paradoxical. Different generations within a speech community at any given time share a language, yet speakers from different periods may speak varieties that are mutually unintelligible. Despite unbroken transmission and acquisition, the language changes. There are two major explanations for this. First, humans can robustly produce and comprehend a wide range of variation, which makes it possible for the language spoken by a community to drift subtly over time. Second, even if intergenerational transmission is constant, it may be discontinuous.

These two factors in linguistic change have been recognized since at least Baudouin de Courtenay (1972) and Saussure (1959) and were given theoretical precision by Kiparsky (1988; 1995) and Bermúdez-Otero (2007; 2014a) among others. Discontinuous learning is made possible by recognizing that learners compile a mental phonological system from the physical phonetic utterances of their parents. That is, parents are assumed to know a grammar with the modular, feed-forward architecture common in generative linguistics. The phonological module compiles discrete symbolic representations, which are fed to a phonetic module, where gradient processes implement them as physical sounds.

Diachronic sound change starts as drift in the phonetic component. Gradient phonetic processes can be language-specific (Keating 1985), and thus can become more or less exaggerated over time. If they become sufficiently extreme, then language learners may misperceive them as categorical phonology, not gradient phonetics (Hyman 1976, Ohala 1989; 1992; 1993, see also Blevins 2004, Fruehwald 2013). Learners must then posit a phonological analysis where their parents had only a

phonetic analysis. This process is often referred to as phonologization.

Discontinuous learning occurs when the phonetic system of a speech community crosses a threshold and learners categorize it as part of a novel phonological system. As a schematic example, consider a fairly recent change found in many English dialects. In the seventeenth century, [r] weakened in coda position, and by the late eighteenth century, deletion of coda [r] was well attested (Bailey 1996:98-109, McMahon 2000:234-240, see also Hay and Sudbury 2005 for a sociolinguistic study of New Zealand English). Hence, *manner* varied between [mænər] and [mænə], while *Anna* was pronounced [ænə]. Coda deletion then became so prevalent that the final syllables of *manner* and *Anna* were only distinguished when the [r] was syllabified as an onset ("linking [r]"). At this stage, *manner is* [mænər ɪz] contrasted with *Anna is* [ænə ɪz]. Then, [r] started to appear in words that historically never had it ("intrusive [r]"), as in *Anna[r] is* [ænər ɪz]. The change is summarized below.

(49)

manner	*manner is*	*Anna*	*Anna is*	
mænə	mænər ɪz	ænə	ænə ɪz	Linking [r]
mænə	mænər ɪz	ænə	ænər ɪz	Linking and Intrusive [r]

Though the analysis of [r]-sandhi in modern English is controversial (see McCarthy 1991; 1993, Harris 1994, McMahon 2000, Bermúdez-Otero 2011, Sóskuthy 2013), the most important (and uncontroversial) fact is that a significant change has taken place. Either final [r] has been added after non-high vowels throughout the lexicon, or an epenthesis process has emerged.

Bermúdez-Otero and Hogg (2003:99 ff) discuss how frequency asymmetries may have permitted learners to maintain an epenthesis analysis instead of guiding them to the veridical deletion system (see also Bermúdez-Otero 2011). At the point that deletion became prevalent, learners perceived a categorical alternation between [r] and ∅, and began to construct an appropriate lexicon and a phonological grammar. It would not be unreasonable for learners to analyze the alternation as being the result of epenthesis, so that underlying representations could lack syllable-final /r/. Importantly, prior sound changes had skewed the lexicon so that words like *Anna* that did not alternate were comparatively rare. An earlier apocope process had decreased the number of words without an alternation after [ə] (Minkova 1991). Meanwhile a process known as "pre-[r] breaking"

had greatly increased the number of words that had an alternation after [ə] (see McMahon 2000). With only infrequent counter-evidence against epenthesis from words like *Anna*, a substantial proportion of learners kept it. The rise of intrusive [r] from the linking [r] system was simply the productive application of this grammar.

4.1.1 Change by Design

English [r]-sandhi might be called an "accidental change", since the proposed cause depends heavily on idiosyncratic facts of the English lexicon. Whenever multiple analyses are theoretically available, or a more accurate analysis is ignored, such an accidental cause must feature. Commonly discussed causes in this vein include sensitivity to particular "triggering" data (Lightfoot 1999), the order or age in which data is encountered (Kiparsky 1978; 1995), the relative frequency of structures (Niyogi 2006), or prior biases against particular analyses (Moreton 2008, White 2013; 2014).

A stronger claim is that a particular change follows from the definitional principles of language itself. Such a change might be called a "change by design" Arguments for change by design posit that learners were exposed to data that did not conform to the requirements for human language. This is not as outlandish as it sounds, since it is possible for phonetic drift to move the surface forms produced by a valid human language over categorical thresholds, triggering a novel phonological analysis. All that is required for learners to encounter "non-human" surface forms is for phonetic drift to move a language to the point where the perceived phonological system is impossible. In section 4.6 we will review one such proposal by Albright (2002), but will eventually settle on using the typology defined by sets of Classic OT constraints as the means to delineate possible from impossible languages.

Allowing learners to enforce change by design moves our theory of learning further from the standard task in computational learnability, which is to determine whether a class of languages L (subsets of Σ^*) can be identified by a learning function $\phi()$, see chapter 2. Intuitively, all the languages in the class bear common properties, and when the learning function $\phi()$ is exposed to one language in the class, it exploits these properties to identify which language the observed

69

strings come from. More concretely, a human child may be presumed to be born with or quickly develop some idea of what a human language is, and leverage this knowledge to determine what language(s) are spoken in its environment.[1] In the context of OT, as chapter 2 showed, learners are assumed to learn or already know a complete inventory of markedness and faithfulness constraints, the permutations of which determine a typology of input-output functions. Because multiple rankings of OT constraints may be consistent with the surface forms and alternations of a language, the learning function $\phi()$ must rule out the input-output functions that are contradicted by the observed data.

If a learner is to carry out change by design, however, it is not enough to identify the languages in the typology defined by a constraint set. The learning function $\phi()$ must also recognize when a language is outside of the available typology and respond by producing a grammar for a language that is within the available typology. This chapter will review how a learner can recognize when a language is not representable by any ranking of OT constraints (section 4.2) and section 4.6 extends a proposal developed by Albright (2002) to specify how the learner should respond in such a situation.

4.2   Detecting Non-OT Languages

In OT, the typology of available languages is obtained by permuting the ranking of the assumed constraints. As we have seen in chapter 2, learning from paradigm-labeled surface forms is carried out by accumulating sets of partial rankings, where each set represents a distinct hypothesis to explain the surface inventory and alternations in the data. Because the learner assumed the structure of OT, all sets were internally consistent (i.e. no set had contradictory rankings), while each datum in the observed corpus was the unique optimum of an input for any total ranking that respects the partial rankings in the set. If no ranking of OT constraints is compatible with the observed data, it is because either of of these properties is violated. That is, if no set ensures that each datum is a

---

[1]To assume otherwise is to assume that human children can learn arbitrary subsets of Σ^*. It would be a mistake to adopt such an assumption, not least because of Gold's (1967) proof that it is impossible to learn the relatively unconstrained set of recursively enumerable languages (languages whose legal utterances can be recognized by a finite device recursively applying a finite set of rules) or indeed any strict superset of the finite languages (languages whose legal utterances can be enumerated in a finite list).

unique optimum, or every set is inconsistent with an ERC from a datum, then the learner $\phi()$ has been signaled that the observed language is outside of the assumed typology. See section 2.2.1 for a discussion of how inconsistency is detected.

The only point that must be added is how to detect that the observed datum is not a unique optimum. This is a straightforward question. When computing the contenders of an input (the finite set of non-harmonically bounded outputs), if one contender performs identically to the intended winner on every constraint, then the observed output cannot be distinguished from the contender. As an ERC, this situation is signaled by a vector consisting only of e, as shown below.

(50)

natibadim	*CCC	*V	Max-V
a. ☞ natbadim		***	*
b. ☞ natibdim	e	*** e	* e
c. natibadim	e	**** W	L
d. natbdim	* W	** L	** W

Assuming an input of /natibadim/ for an observed winner *natbadim* with the constraint set in 50 means that there is no way to distinguish between the intended winner *natbadim* and *natibdim*. Under any ranking of these constraints, the observed winner is not the unique optimum for this input. The learner must simply be able to detect a tie to notice that this assumed property of an OT language has been violated. A learner will have to either try a different UR for this output, or if no other UR is feasible, announce that the observed language is not in the assumed typology. See also Tesar and Smolensky (1993, §4.1).

True ties are generally not considered to be a realistic problem in OT, since there is a large enough inventory of postulated constraints that some constraint, however tangentially related to the phenomenon being analyzed, will distinguish between any pair of candidates. However, in chapter 5 we will consider a language where tied candidates on relevant constraints are so common

that recruiting tangentially related constraints to consistently pick winners is infeasible, so tied candidates will be expositionally relevant.

4.3   Alternations (Partly) Cause Inconsistency

The learning system discussed here uses OT to define the space of possible languages. Importantly, of the two targets of phonological learning, the phonotactic distribution cannot by itself place a language outside of the space of OT-languages. To see this, assume a set of OT constraints CON containing markedness constraints and faithfulness constraints such that any deviation from a string penalized by a markedness constraint is penalized by at least one faithfulness constraint. Recall from section 2.3 that phonotactic learning seeks to ensure the legality of the corpus of observed surface forms. To acheive this, the learner must rule out grammars that cannot derive utterances in the observed data, that is, the grammars that do not dominate violated markedness constraints with constraints that favor observed forms. Because faithfulness constraints do not inherently disprefer any surface form (unlike markedness constraints), a perpetually available grammar that derives all observed surface forms is one where all markedness constraints are dominated by every faithfulness constraint. Such a grammar may be called the "identity grammar", since it ensures that all forms in any corpus of data are derived from URs that are identical to the observed SRs. The upshot of the availability of the identity grammar is that no matter how many marked structures are present in a corpus of data, there is always a grammar that permits all of them to surface. A phonotactic distribution is always describable by the constraints in CON.

Note that this is not a claim that any set of constraints CON can make the gaps of any phonotactic distribution be principled gaps. For instance, if CON is the set of constraints that define human phonological patterns, it will include markedness constraints that disfavor voiced obstruent codas and will not include markedness constraints that favor them. If a learner is presented with a language that has voiced obstruent codas to the exclusion of voiceless codas, it will learn a ranking that permits voiced codas, but will be unable to ensure that voiceless codas are illegal. The observed surface forms will simply be a subset of the language derived by the grammar. Clearly, the prediction from OT is that if sound change produced a language with solely voiced obstruents

(Blevins 2004, Yu 2004, Kiparsky 2006; 2008), learners would be able to maintain exclusively voiced codas in words they had encountered. However, their grammar would accept voiceless obstruent codas, making the restriction of codas to voiced segments be merely an accident of the lexicon. We would expect speakers of such a language to not modify loan words with voiceless obstruent codas, or to judge nonce words with voiceless obstruent codas as acceptable.[2] See section 4.5 for further discussion of how a learner could respond to a language that violates the assumed typology.

Where any phonotactic distribution is compatible with some OT grammar with constraints in CON, alternation patterns do not benefit from such a guarantee. Beyond the trivial case where CON does not contain markedness constraints that can compel some alternation, it is possible to imagine sets of paradigms that satisfy the same markedness constraint by different faithfulness violations. For instance, complex coda violations could be resolved by MAX violations in [and-a, an], but by DEP violations in [and-a, andi]. If both paradigms were in the same language, a ranking paradox between MAX and DEP would result.

This is not simply a result of picking a convenient example. The nature of morphophonological learning is different from phonotactic learning. Phonotactic learning has a restricted pool of constraints that might disfavor observed forms (markedness constraints), and a separate pool of constraints that can always favor observed forms (faithfulness constraints). This situation does not obtain in morphophonological learning, since alternations show which faithfulness constraints prefer losers. Since the observed forms may still violate markedness constraints, there is no corresponding guarantee that markedness constraints will only prefer winners. In morphophonological learning there is no pool of constraints that can be counted on to ensure consistency.

Morphophonological alternations may not only be internally inconsistent, but can also be inconsistent with the phonotactic distribution. Some ranking that may be necessary to keep observed forms legal could be incompatible with what is demanded by alternations. For example, consider a language with contrastive voicing in all environments, so that *ag*, *aga*, *ak*, *aka* are legal. The necessary constraint ranking is ID-VOI ≫ *VOIOBS#, *VOIOBS, *IVV. Barring underspecification,

─────────────────────

[2]Kiparsky (2006; 2008) argues that there is no evidence that diachronic paths that end in word-final voicing are taken to completion. Given the ability of OT to accomodate any phonotactic distribution, we cannot rely on the architecture of OT to ensure that synchronic voiced codas never arise via sound change.

a paradigm with a voice alternation like [ag-a, ak] demands that ID-VOI be dominated by *IVV or *OBSVOI#. If such a paradigm exists in the language, the morphophonological data will require rankings that are incompatible with rankings required for the phonotactic data.[3]

Indeed, this situation has already been encountered in the previous chapter, where a morphophonological hypothesis could be rejected if it conflicted with the phonotactics. The important point here is that all morphophonological hypotheses may be inconsistent with the phonotactic ranking requirements. In such a case, a learner should arrive at the conclusion that the observed language is outside of the assumed typological space. There are several imaginable next steps for a learner in such a situation, including just giving up, adopting a default language, or seeking a language that resembles the observed language to some degree. Later sections of this chapter will explore the latter possibility.

4.3.1 Opacity and OT

The discussion thus far in section 4.3 focuses on OT languages in the abstract, assuming only that OT languages are total functions between $\Sigma^* \times \Sigma^*$, as defined by the permutations of the constraints in the set CON. The only assumption made about CON was that it contains markedness and faithfulness constraints. Obviously, the content of CON matters quite a bit, as it determines which functions between phonological inputs and outputs are actually possible in any given case. A narrow (and mundane) way for a language to fail to be an OT language is simply if CON does not contain constraints that can be ranked to make the required input-output mapping optimal.

A more substantial cause of OT being unable to generate a particular input-output mapping lies in the key OT tenet of parallel evaluation, whereby all possible outputs are evaluated simultaneously and the best output as determined by the constraint ranking is selected. This contrasts with serial evaluation, where the output of an input is the product of a chain of operations, as occurs in rule based phonology (Chomsky and Halle 1968) or in Harmonic Serialism (McCarthy 2010).

---

[3]It is well known that such systems of voice alternations in fact exist, as in Turkish (Inkelas, Orgun and Zoll 1997), a fact that motivates an underspecification analysis. In general, featural alternations that contradict unambitious phonotactic rankings can be generated by proposing underspecification. However, alternations that do not properly involve features, like insertion/deletion, moraic alternations, or stress alternations are not amenable to underspecification and are thus more prone to contradicting phonotactic requirements.

One of the crucial effects of using parallel evaluation is that phonological phenomena that require reference to a representation that is intermediate between the UR and the SR cannot be generated (see especially McCarthy 2008, and chapter 5). Many of these phenomena that OT cannot generate are what were traditionally described with opaque rule interactions (Kiparsky 1968b; 1971; 1973). There continues to be debate over the proper meaning of opacity in an OT setting (Bakovic 2007; 2011, McCarthy 2007a, Tesar 2013), and a variety of novel constraint types and other modifications to OT have been proposed to augment the range of patterns it can produce. Nonetheless, opaque alternations are widely recognized as problematic for OT, as they very frequently require inconsistent rankings.

The goal of the discussion here is not to provide an argument for or against the validity of opaque phenomena generally. However, with the fomulation of a learning theory, it does attempt to provide some precision to Kiparsky's original thesis that opaque phonology may be difficult to learn and hence will be prone to be changed by learners (see especially section 4.8). The change that our learner will enforce is a paradigmatic change known as levelling, which the next section takes up.

4.4   Responding to Inconsistency

On the face of it, the stated goal of responding to an inconsistent language by producing a maximally similar, but consistent, language is straightforward. Once the corpus contains data such that all hypotheses become inconsistent, the learner could construct all consistent subsets of the ERCs that form the hypotheses, and compare the subsets on an accuracy metric for the observed language. Rather than seek an exact solution to this problem, our algorithm will sacrifice alternations and keep phonotactic generalizations, thereby actuating a paradigmatic change known to historical linguists as levelling (see section 4.4.2).

4.4.1 Why not Sacrifice Phonotactics?

It is not obvious that alternations must be sacrificed when inconsistency results, since one potential reason for inconsistency is a conflict between phonotactics and alternations. Importantly, the historical record contains events dubbed by historical linguists "extension" where phonotactics are overturned. English [r]-sandhi, discussed in section 4.1, is one such case, since before intrusive [r] was generalized from linking [r], the phonotactics permitted non-high vowels in hiatus. The imposition of intrusive [r] in words that historically permitted hiatus amounts to the sacrifice of a phonotactic pattern to an alternation. Other cases of extension include Portuguese laxing (De Chene 2010) and an ongoing change in Korean (Albright and Kang 2008).

Frequency features prominently in prior explanations for why a phonotactic pattern might be overruled by an alternation. As mentioned in section 4.1, the English lexicon had been heavily skewed so that alternation in hiatus was more frequent than non-alternation, while Albright and Kang (2008) point out that the pivot forms in the Korean change are dramatically more frequent in child-directed speech than the changed forms.[4] While the precise role of frequency in change is still to be worked out (see Albright and Kang 2008), it seems probable that some frequency statistic may help to decide whether alternations or phonotactics are sacrificed. At the very least, this question will not be immediately relevant for us, since the case studies considered in section 4.8 and chapter 5 all involve the sacrifice of alternations for phonotactics.

4.4.2 Levelling Introduced

Leveling is usually diagnosed observationally by one allomorph of a morpheme (the pivot) gaining a wider distribution by replacing other allomorphs of the morpheme. A preliminary example of this will appear in section 4.5, while section 4.6 will refine this idea further, The classic case of levelling is the Latin *honor* analogy (Kiparsky 1971, Kenstowicz 1996, Albright 2002; 2005), Gorman 2012), whereby the Classical Latin paradigm for 'honor' shows the generalization of what

---

[4]The English change also featured a more frequent pivot, since word-final consonants are three times more likely to be parsed into a coda than resyllabified into an onset in English (Bybee 1985:73). The most frequent allomorph in paradigms that had an [r]-∅ alternation was the [r]-less allomorph, possibly leading human learners to the conclusion that the less frequent allomorphs had arisen via epenthesis.

was originally an oblique case allomorph in Old Latin, as shown below:

(51)

	Old Latin		Classical Latin
nom.sg	hnoːs	>	honor
gen.sg	honoːris	>	honoːris

The description of levelling as involving a "pivot" allomorph that "spreads" to other cells of the paradigm has no formal status in the theory developed here. Any change that is wrought within a paradigm occurs via a change in the UR for the morpheme in question, which the grammar, due to the high rank of faithfulness constraints induced during phonotactic learning, realizes faithfully in new contexts. There are other commonly attested historical outcomes for paradigms as well, including the memorization of an irregular paradigm, the development of new lexical entries, and the maintenance of an irregular form with special semantics alongside a phonologically regular form with regular semantics. Such changes are not of direct interest here, though a full explanation of language change would clearly explain them as well.

4.5 Schematized Impossibility

As an example of how sound change might draw a language into inconsistency, consider a language that degeminates consonants word-finally and subsequently lenites geminate consonants to voiceless consonants and singleton consonants to voiced consonants, as in (52).

(52)

'Horse'		'Bridle'			
acc	nom	acc	nom	unattested	Stage
at-a	at	att-a	at	*att	Degemination
ad-a	ad	at-a	ad	*at	Lenition

As Kiparsky (2008) has pointed out, one might expect a learner that was exposed to surface forms in the degemination stage in (52) to acquire a grammar that correctly produced the observed data. The expectation of successful acquisition is based on the widely observed and uncontroversial stability of the pattern of degemination (see Kaye and Nykiel 1979:76-79, and Kennedy

77

2003:80-85). Furthermore, it would be unsurprising if speakers with a degemination grammar developed a phonetic sound change leniting geminate and singleton consonants to voiceless and voiced segments. However, what learners would do in response to the phonetic lenition data is controversial.

Much research argues that phonology is phonetically natural (Stampe 1973, Hayes 1999a, Hayes, Kirchner and Steriade 2004, White 2013). Hayes (1999a,§6.2) summarizes several phonetic reasons to expect voicelessness to be favored over voicing in word-final position, while there are not phonetic factors favoring the reverse. The lenition stage in (52) represents a language where not only are voiceless segments absent word-finally, but an alternation actively enforces voicing at the end of the word. If CON does not contain constraints that favor voicing word-finally, then the perceived language at the lenition stage is outside of the available typology for learners. Once the learner detects inconsistency, following the idealized response to inconsistency adopted here, the phonotactic ranking will force the alternating paradigm to lose its alternation.

Under the standard theory of CON adopted by phonologists, the only constraint regulating voice in word-final position, *VOIOBS#, makes voicing marked in this context. In order for forms like *ad* to be legal, such a constraint must be dominated. Indeed, since the claim is that no markedness constraint prefers word-final voiced obstruents to other segments, the phonotactic ranking requires that *VOIOBS# be dominated by a faithfulness constraint like ID-VOI:

(53)

ad	*VOIOBS#	ID-VOI
a. ☞ ad	*	
b. at	L	*W

Importantly, this constraint set lacks markedness constraints that can force the alternation observed between *at-a* 'bridle-acc' and *ad* 'bridle'. No matter what UR is chosen for the morpheme 'bridle', all allomorphs of the morpheme cannot be generated in the appropriate contexts. For instance, if /at/ is chosen, *at-a* is generated (as it violates no constraints):

(54)

at-a	*VOIOBS#	ID-VOI
a. ☞ ata		

But when it comes time to generate the form *ad*, the desired winner (marked with a frowning face) isn't even a contender. An intended loser wins instead. This morphophonological hypothesis is not only inconsistent with the phonotactic ranking, it is inconsistent outright.

(55)

at	*VOIOBS#	ID-VOI
a. ☹ ad		*
b. ☞ at		L

Likewise, the alternative interpretation of the alternation as devoicing fails, since *at-a* is harmonically bounded from /ad-a/:[5]

(56)

ad-a	*VOIOBS#	ID-VOI
a. ☹ ata		*
b. ☞ ada		L

With every morphophonological hypothesis being inconsistent, the learner has been signaled that the observed language is outside of the available typology. Following the mandate set in

---

[5]Adding a constraint that forces devoicing, or fortition, in onset position (as advocated for Lezgian by Kiparsky 2008) would allow the intended winner *at-a* to be a contender. However, the ERC that allows *at-a* to beat *ad-a* would be inconsistent with the phonotactic rankings, which would allow *ad-a* as a legal form, due to it being attested as the accusative of 'horse' in (52). Kiparsky's analysis of Lezgian does not encounter this same problem, as there is no opposing non-alternating voiced segment in the surface inventory of the language.

section 4.4, the learner will discard the inconsistent morphophonological ERCs but retain the consistent phonotactic ERC set. Fortunately in this example, the phonotactic ERCs provide a total ranking of the constraint set, so that no further elaboration must be performed (see section 4.5.1 immediately below). With the grammar set, the final realization of the observed paradigm *ad, at-a* will depend only on what the UR is set to. If it is set to /ad/, then the grammar will produce *ad, ad-a*, effecting a levelling where the nominative is the pivot. While if it is set to /at/, the grammar will produce *at, at-a*, where the accusative allomorph appears to spread its distribution.

4.5.1 Elaboration of Rankings Post-Inconsistency

If the full language requires an inconsistent ranking, the strategy of the learner must shift. The earlier strategy of gathering ERCs from each individual datum was based on the assumption that there was a correct grammar that could generate all surface forms while deriving paradigmatic alternations. Once inconsistency has become certain, rankings can no longer be distinguished along the lines of possibly correct versus definitely incorrect, but instead are just more or less incorrect.

Recall that it is possible to address phonological learning either by enumerating every total ranking of constraints or by the more common method of gradually accumulating ranking statements (see chapter 2). These options are still available once alternations have been found to render the ranking inconsistent, though the enumeration strategy only requires listing the rankings that are compatible with the phonotactic ERCs, instead of listing all total rankings of constraints. A major perceived benefit of the accumulation strategy over the enumeration strategy is speed, as direct consideration of every possible grammar is thought to be too slow to plausibly model human language acquisition (see for instance Tesar and Prince 2007). However, the speed advantage of the accumulation approach goes away once grammars are only more or less correct. Magri (2013) has shown that the problem of finding the OT grammar that maximizes accuracy is NP-complete, meaning that there is no known way to quickly solve it. At the time of writing, the slowness argument against an enumeration strategy is inapplicable.

There is also an important obstacle to accumulating ERCs in a language where a full analysis

would require an inconsistent ranking. It is possible that an observed allomorph is harmonically bounded under every underlying specification, or at the very least every underlying specification for some morpheme requires a ranking that is contradicted elsewhere. The observed allomorph clearly cannot win under any consistent ranking, and if some consistent ranking is to be found, the allomorph must not be designated as the winner. However, the learner must then decide which unobserved candidate to declare as the winner. Such a problem may not be insurmountable, as the learner could compare potential winners to the observed allomorph for phonetic similarity, but it is a decision where the correct choice is not obviously available. See section 5.7.3.1 for some discussion of gradient comparison of generated winners to observed allomorphs. If a total ranking is assumed on the other hand, the winner is determined by the ranking, and the learner simply needs to check whether the winner matches the observed allomorph.

If we assume that the learner's goal is to maximize accuracy even when perfect accuracy is impossible, enumeration of all of the remaining hypotheses is actually a rather straightforward way to proceed. Because we are assuming that phonotactic ERCs are kept and the morphophonological ERCs are discarded, accuracy only needs to be calculated for alternations, not whether any surface forms are rendered illegal. To tabulate accuracy, the learner could follow the following procedure. For each morpheme in the corpus, the learner can obtain the URs that would generate its allomorphs in their observed environments via the comprehension methods discussed in Eisner (2002) or Riggle (2004:194 ff).[6] The sets of URs for each allomorph would then be intersected with each other. Each time a non-empty intersection is discovered, an alternation is generated by the grammar, and the accuracy score may increase. Once this score has been compiled for all rankings, the grammar(s) with the highest score should be selected. This approach will be modified slightly after the discussion in the next section, after which it will be summarized in pseudo-code.

---

[6]This is actually an advantage of the enumeration approach, since the comprehension methods presuppose a total ranking.

4.6 The Single Surface Base Hypothesis

A consistent observation is that paradigm levelling within a language consistently spreads an allomorph from the same morphological or phonological environment in all paradigms (see for instance Garrett 2008, §1).[7] Hence, in addition to a mechanism to assess overall accuracy, it is also prudent to specify that allomorphs that occur in a particular morphological or phonological environment are in some sense privileged. Many proposals to ensure this consistency in levelling exist, but we will here focus on an especially influential model, the Single Surface Base hypothesis (Albright 2002; 2005; 2010). This section will review key elements of the Single Surface Base hypothesis, section 4.7 discusses how the key insights of the Single Surface Base hypothesis can be incorporated into the model proposed here, and section 4.8 will discuss the case of Yiddish paradigm leveling in the context of the proposals developed here.

In its original form, the Single Surface Base hypothesis makes a concretist claim about phonological structure: the allomorphs of any morpheme are derived from some allomorph of the same morpheme. Put another way, the Single Surface Base hypothesis restricts the inputs of phonology to be a subset of its outputs, or in OT terms, IO-faithfulness constraints are indistinguishable from OO-faithfulness constraints (Benua 1997, Burzio 1998; 2002). The hypothesis gets its name from the stronger restriction it imposes: the allomorphs of morphemes in a morphological category, like nouns or non-past verbs, are derived from the allomorph occurring in the context of a particular morphological feature, like genitive or first person.[8] For convenience, rather than refer to "allomorph occurring in the context of a particular morphological feature", I will refer to a "paradigmatic cell", although a precise definition of morphological paradigms will not be sought here.

To see the practical implications of these restrictions, consider the following synthetic example adapted slightly from Albright (2002:46), featuring a language with palatalization of [k] before [i].

---

[7]This is evidently only a very robust tendency, as there are cases where allomorphs from different morphological environments have been spread (see Winter 1971:59-60).

[8]Affixes are themselves instantiations of morphological categories, for instance, number, case, or aspect. It is conceivable that affixes are derived from a particular morphological context, as in number affixes being derived from the allomorph that appears in the context of a locative suffix. However, the treatment of affixes remains unsettled, though Albright and Kang (2008) tentatively suggest grouping affixes by phonological properties.

(57)

	Absolutive	Ergative
1.	ʔak	ʔa͡tʃ-i
2.	muk	mu͡tʃ-i
3.	lok	lo͡tʃ-i
4.	sa͡tʃ	sa͡tʃ-i
5.	ru͡tʃ	ru͡tʃ-i
6.	da͡tʃ	da͡tʃ-i
7.	lot	lot-i
8.	ɡup	ɡup-i
9.	lap	lap-i
10.	ban	ban-i
11.	yul	yul-i

The Single Surface Base hypothesis provides a ready answer to the question of which allo-morph should be preserved and spread during leveling. At the very least, this allomorph is the base allomorph. No matter if the ergative or absolutive cell is chosen as the base, the base form will always be correctly generated within its original environment, since it can be perfectly predicted from itself.

Importantly, however, while the model stipulates that only a surface form from a particular context be the base, which context is chosen is determined by analyzing the overall predictability of the non-base forms from the base forms. In the case of a language like that in (57), the distinction between non-alternating [t͡ʃ] and [k] is erased before the suffix [-i]. If the absolutive cell is selected as the base, it is straightforward to elaborate the phonotactic grammar with a general process that palatalizes [k] to [t͡ʃ], while leaving [t͡ʃ] unchanged. By contrast, if the ergative case is selected as the base, then there are limited options for elaborating the phonotactic grammar, since generating the absolutive forms of *ru͡tʃ-i* (*ru͡tʃ*) and *mu͡tʃ-i* (*muk*) will require either a lexically specific rule depalatalizing [t͡ʃ] in the environment [mu_#], or there will be no way to predict whether the absolutive form ends in [k] or [t͡ʃ].

Albright (2002; 2005) discusses a variety of statistics to quantify which cell of the paradigm

best supports the alternations in the data. The statistics are primarily dependent on the representations produced by the Minimal Generalization Learner (Albright and Hayes 2002). A more general statistic for this question is conditional entropy, as is discussed by Ackerman, Blevins and Malouf (2009). To my knowledge, the literature has not yet determined which of these statistics is more appropriate.

Note that phonological predictability is not the only proposed guide for how to decide what which form to privilege in levelling. The literature on levelling has suggested at various times that the most frequently occuring allomorph, the allomorph occurring with the fewest affixes, the least marked morphological features, the least phonologically marked structures, and others may all be important criteria. Note though, that Albright (2010) discusses Yiddish paradigmatic change where all of these criteria fail to correctly predict the basic cell. Ultimately for our purposes, it matters only that there be *some* way to determine which allomorph to privilege, and our discussion will not decide between these competing metrics.

4.7 Reinterpreting The Single Surface Base

Despite its utility, the original formulation of the Single Surface Base hypothesis clashes with the results from elsewhere in this dissertation. Chapters 2 and 3 argued that there is stable human phonology that is best described with composite URs, i.e. the set of inputs is not limited to be a subset of outputs. Chapter 5 argues that composite URs also appear in phonological systems where inconsistency was detected and leveling took place. In terms of the theory developed here, the evidence presented in chapter 5 indicates that even after inconsistency has been detected, the central tenet of the Single Surface Base hypothesis is violated.

The insights of the Single Surface Base hypothesis can be reconciled with a model that permits composite URs if the role of a single surface allomorph is reinterpreted. Where the original formulation of the Single Surface Base hypothesis posits that all allomorphs of a morpheme are derived from a single paradigmatic cell, the single paradigmatic cell can be recast as a privileged output. Under the original formulation the goal of learning was to construct a grammar that derived other allomorphs from the base allomorph, with this reinterpretation the goal is to construct a gram-

mar that derives other allomorphs while ensuring that the base allomorph surfaces. Nothing else need change in this reinterpretation. Most saliently, the requirement that the privileged allomorph always be drawn from the same paradigmatic cell remains unchanged.

This reinterpretation of the Single Surface Base hypothesis is easily incorporated into our theory. Rather than counting any generated alternation towards the accuracy count proposed in section (4.5.1), only alternations that can be generated from the URs for the privileged output are counted. Needless to say, the entire paradigm will also be generated from the URs for the privileged output. With the UR for the paradigm yoked to the privileged output, the descriptive generalization that the privileged output is both stable in cases of levelling and spreads to other cells will be ensured by the high rank of faithfulness constraints left over from phonotactic learning. The procedure that is followed once inconsistency has been detected is summarized in the following pseudo-code.

In prose, the algorithm in 3 carries out the following procedure. The procedure starts with an enumeration of rankings R (presumably only those consistent with the set of phonotactic ERCs), a text of paradigm labeled surface forms T and a specification of the morphological or phonological context in which the privileged allomorph occurs.[9] Each ranking in R is associated with a with a score, which is contained in S. The algorithm then loops through the set of rankings, augmenting the score of a ranking every time the set of URs that can be mapped to the privileged allomorph overlaps with the set of URs that can be mapped to a non-privileged allomorphs. The algorithm rewards overlapping sets of URs, because an overlapping set of URs means that under the current ranking, an observed alternation will still be generated. Upon completion of the loop, the algorithm returns the rankings that achieved the highest score.

4.8   Yiddish Levelling

Under the original formulation of the Single Surface Base hypothesis, neutralizations in the base cell are the root cause of paradigmatic change. As discussed above, if the toy language in (57) were to have the ergative cell be specified as the base, the collapse between the categories [k] and [t͡ʃ]

---

[9]There is a significant imprecision in the pseudocode. First, an undefined function *alreadyObserved()* is used on line 21. The intended meaning is that the function determines if an allomorph of a morpheme has already been observed.

Algorithm 3 Response to Inconsistency

1: **function** MOSTACCURATE(R, T, C) ▷ Rankings = R, text = T, privileged context = C

2: $n \leftarrow length(R)$

3: $S \leftarrow \langle 0_1, 0_2 \dots 0_n \rangle$

4: **for** $i \in range(n)$ **do**

5:   $r = R[i]$

6:   $A \leftarrow \emptyset$ ▷ Non-privileged allomorphs

7:   $ObA \leftarrow \emptyset$ ▷ Observed morphemes (non-privileged)

8:   $P \leftarrow \emptyset$ ▷ Privileged allomorphs

9:   $ObP \leftarrow \emptyset$ ▷ Observed morphemes (privileged)

10:   **for** $f \in T$ **do** ▷ $f = \langle string, \langle morphemes \rangle, \langle allomorph\ indices \rangle \rangle$ (section 2.1.2)

11:    **for** $j \in range(length(form[1]))$ **do**

12:     $m \leftarrow form[1][j]$

13:     $a \leftarrow f[0][f[2][j]]$

14:     $URs \leftarrow comp(a, f, r)$ ▷ See Eisner (2002), Riggle (2004:194 ff)

15:     **if** $context(a, f) \in C$ **and** $m \notin ObP$ **then**

16:      $P \leftarrow P \cup \{\langle m, a, URs \rangle\}$

17:      $ObP \leftarrow ObP \cup \{m\}$

18:      **if** $m \in ObA$ **then**

19:       **for** $b \in A$ **do**

20:        **if** $b[0] = m$ **and** $b[2] \cap URs \neq \emptyset$ **then** $S[i]+ = 1$

21:     **else if** $context(a, f) \notin C$ **and** $alreadyObserved(a, m) =$ False **then**

22:      $A \leftarrow A \cup \{\langle m, a, URs \rangle\}$

23:      **if** $m \in ObP$ **then**

24:       **for** $p \in P$ **do**

25:        **if** $p[0] = m$ **then**

26:         **if** $p[2] \cap URs \neq \emptyset$ **then** $S[i]+ = 1$

27:         **break**

28: **return** $\langle R[j]$ if $R[j] = max(S) \rangle$

---

before [i] would pose problems for generating the absolute forms, and either non-alternating [t͡ʃ] would begin to alternate with [k], or alternating [t͡ʃ] would stop alternating. The position taken here is that neutralizations are not by themselves a hurdle for learner, rather neutralizing alternations that cannot be generated by the grammar are the ones that must be changed. As it turns out, attested paradigmatic changes that have been adduced in arguments supporting the Single Surface Base hypothesis often involve not just neutralizing alternations, but opaque neutralizing alternations. This section reviews how our model explains one such case study, focusing on the levelling of vowel length and consonant voice alternations in Yiddish (Albright 2004; 2008; 2010).

4.8.1 Precursor to Levelling

The Yiddish levelling we are concerned with has its roots in an opaque system that emerged in Middle High German, which is often taken to be roughly the last common ancestor language for Yiddish and modern German. In the 14th century, Middle High German (MHG) developed word-final devoicing and, to a first approximation, open syllable lengthening. This was followed by the innovation of word-final schwa deletion (apocope), which rendered the first two processes opaque. Crucially, both nouns and non-past verbs had a suffix [-ə] that was targeted by the apocope process, resulting in opaque alternations, as shown in (58).

(58)	'say-1.sg'	'say-2.sg'	'praise'	'praise-nom.pl'	
	/sag-ə/	/sag-st/	/lob/	/lob-ə/	UR
	—	—	lop	—	Devoicing
	saːgə	—	—	loːbə	Open σ Lengthening
	saːg	—	—	loːb	Schwa Apocope
	[saːg]	[sagst]	[lop]	[loːb]	

The deletion of word-final schwas meant that word-final devoicing underapplied in forms like *saːg* 'say.1.sg' and *loːb* 'praise.nom.pl'. Under the current description, we also have that open syllable lengthening overapplied in the same forms. However, there is some uncertainty regarding the data, as King (1988) reports that monosyllabic nouns in fact had already gained long vowels in all allomorphs, with only one known exception. Hence, it seems likely that the paradigm for 'praise'

before schwa apocope was innovated had the nominative singular form *loːp* and the nominative plural form *loːb-ə*. Multisyllabic nouns are not mentioned by King, but presumably length had not spread to all members of the paradigm.

4.8.2 Aftermath of Opacity

The opacity wrought by apocope meant that obstruent voice in word-final position was no longer predictable in surface forms, and that vowel length was similarly not predictable by syllable type.[10] The situation was made worse by the fact that morphemes had alternations in these features, causing each allomorph to need its own UR. Subsequently, the alternations were lost. It is hard to pinpoint exactly when and how the alternations were lost, but the modern Yiddish paradigms for the forms in (58) are *zɔg* 'say.1.sg' and *zɔk-st* 'say-2.sg', and *lɔ͡ɪb* 'praise' and *lɔ͡ɪb-n̩* 'praise-pl'.[11] The crucial point to observe is that [ɔ] is the Yiddish reflex of Middle High German long [aː], while [ɔ͡ɪ] is the Yiddish reflex of Middle High German long [oː]. Hence, it appears that during the development of Yiddish, vowel length and in the case of nouns, obstruent voice as well, spread from one part of the paradigm, as depicted in (59).

(59)

MHG		Pre-Yiddish		Yiddish	
lop	>	*loːb	>	lɔɪb	'praise'
loːb	>	*loːb	>	lɔɪb-n̩	'praise-pl'
sag-st	>	*saːg-st	>	zɔk-st	'say-2.sg'
saːg	>	*saːg	>	zɔg	'say.1.sg'

The alternations from Middle High German may also have been severely weakened in the course of the development of modern German. King (1988) states that the rules governing vowel length in modern German have been lost, or are at best minor processes. Strikingly, Gress-Wright (2010) reports that even though modern German has word-final devoicing, the spelling of devoicing stopped soon after the advent of apocope. Whether this was due to a change in spelling norms or

---

[10]Even before apocope was innovated, vowel length did not correspond perfectly with syllable type (King 1988:28-29, Albright 2010:491). For instance, a degemination process had resulted in words containing V̌CCV becoming V̌CV.

[11]See Albright (2010:520-521) for the realization of suffixes romanized as *-en* in modern Yiddish.

due to the actual loss of word-final devoicing (prior to re-emerging in the modern language) is a matter of some dispute.[12]

That alternations would be lost after they were opacated has led several authors to speculate that there is a causal link between opacity and paradigmatic change (Kiparsky 1968b, Vennemann 1972, King 1976). As the next section shows, the model of learning proposed in this dissertation enforces paradigmatic change in response to opacity.

4.8.3 Actuating Yiddish Levelling

The critical point in the development of Middle High German for our purposes was when what was presumably a phonetic process of word-final schwa reduction became misperceived as deletion. For the sake of argument, assume that the surface lexicon was simply *lop* 'praise.nom.sg' and *loːb* 'praise.nom.pl', so that both vowel alternations and consonant alternations can be addressed in the same paradigm. Assume also the following constraint set:

(60)   a. *VːC]$_\sigma$: Assign one violation for every long vowel in a closed syllable.

     b. *V̆]$_\sigma$: Assign one violation for every short vowel in an open syllable.

     c. *VOIOBS#: Assign one violation for every word-final voiced obstruent.

     d. ID-VOI: Assign one violation for every instance of the feature VOICE that differs between input and output.

     e. ID-LEN: Assign one violation for every instance of the feature LONG that differs between input and output.

As is by now familiar, the learner starts with two goals: ensure the legality of the observed surface forms and compel the observed alternations. In order to ensure that the observed surface forms are legal, the faithfulness constraints ID-VOI and ID-LEN must have a fairly high rank. For instance, when the learner encounters *loːb* 'praise.nom.pl', the major markedness constraints that would enforce word-final devoicing and vowel length regularities must be dominated, as shown in (61).

---

[12]King (1976:5) states that the Yiddish from communities with strong Polish contact re-asserted devoicing as well.

(61)

	lo:b	*V:C]$_\sigma$	*V̆]$_\sigma$	*VoiObs#	Id-Len	Id-Voi
a. ☞	lo:b	*		*		
b.	lo:p	*		L		*W
c.	lob	L		*	*W	
d.	lop	L		L	*W	*W

Once, *lop* 'praise.nom.sg' is encountered, it becomes clear that the grammar needs to compel a faithfulness violation in both voicing and vowel length. However, no matter which direction the unfaithfulness goes in, the required rankings are incompatible with the high rank of faithfulness constraints. For instance, if *lop* were to carry all the unfaithfulness from a UR like /lo:b/, every ranking necessary for *lo:b* to be a legal word must be overturned, as shown in (62).

(62)

	lo:b	*V:C]$_\sigma$	*V̆]$_\sigma$	*VoiObs#	Id-Len	Id-Voi
a. ☞	lop				*	*
b.	lo:b	*W		*W	L	L
c.	lo:p	*W			L	*
d.	lob	W		*	*	L

Examining the constraint set and the rankings required for the observed surface forms to be legal make further examples unnecessary. The markedness constraints that governed the formerly lawful distributions of obstruent voice and vowel length must be demoted below faithfulness constraints. Since faithfulness constraints prohibit change from the underlying specification, theiir

90

high rank prohibits alternations. The learner cannot successfully learn the ambient language.

At this point, as this chapter has discussed, the learner's goals shift away from capturing all observed unfaithfulness. The question turns instead to which grammar, if any, can generate alternations in addition to the privileged allomorph. Following Albright's argumentation, the privileged allomorph was the plural in nouns (and the first person singular in verbs). With this final point in place, the learner will produce a paradigm that matches the historical change. To see this, note first that with identity URs high in the phonotactic ranking, every total ranking that respects the phonotactic ERCs will produce the UR /loːb/ for the privileged allomorph *loːb*. As shown in (61), the phonotactic rankings require that this UR be mapped to *loːb*. Hence, the observed nominative singular form *lop* will be replaced by *loːb*.

To summarize the discussion of Yiddish, after apocope opacated open syllable lengthening and word-final devoicing, our learner produces the following analysis. The final ranking is ID-LEN $\gg$ VːC]$_\sigma$ and ID-VOI $\gg$ *VOIOBS#. The UR for 'praise' is /loːb/, and the resulting paradigm has invariant length and consonant voicing.

4.9   Local Summary

This chapter has proposed that when no constraint ranking can satisfy all of the requirements posed by the data, the learner retrenches to find a ranking that satisfies most of the requirements. Inconsistency is often the result for OT when phonology is opaque, and we have reviewed part of the history of Yiddish, a case of paradigm levelling evidently triggered by opaque phonology. Importantly, the Latin *honor* analogy also was a response to opacity, as the $s \rightarrow r$ process was counterfed by degemination and a large influx of foreign words. See Kiparsky (1971), Kenstowicz (1996), Albright (2002; 2005) and Gorman (2012) for further discussion of the *honor* analogy.

CHAPTER 5

Odawa: Composite URs and Levelling

5.1 Introduction

Having developed a theory for how learners can force paradigm levelling, this chapter turns to a case of dramatic restructuring that took place in the early 20th century in the Odawa and Eastern dialects of Ojibwe (Algonquian, United States and Canada). The change was rather intricate, so we will take some time to lay out the particulars before showing how our model derives the change in section 5.7. Specifically, a phonetic process of vowel reduction applying to stressless syllables in left-to-right iambic feet became so extreme that it was confusable with categorical deletion (Bloomfield 1957, Kaye 1973; 1974b, Piggott 1980). We will we refer to categorical deletion in this configuration as rhythmic syncope.[1] The next generation, however, did not phonologize the pattern, undertaking instead a massive reanalysis of the language.

The key aspect of the reanalyzed language is that the classic rhythmic syncope alternations have been removed by paradigm leveling (Rhodes 1985a, Valentine 2001). This is not to say that some speakers cannot produce some, or even many, paradigms with classic rhythmic syncope alternations. Indeed, it would be surprising if all forms from such a recent stage of the language were completely absent. But the productive pattern, as recorded in nearly every entry in Rhodes (1985a) and substantiated by my own fieldwork at Walpole Island, is the use of levelled paradigms. Native speakers of the language recognize that "you can never go wrong" using levelled forms, even while more experienced or conservative speakers retain some non-levelled forms (Corbiere Corbiere:21, quoted in Valentine 2001:66).

---

[1]The term "rhythmic syncope" is borrowed from Kager (1997). In McCarthy (2008), this process is referred to as "metrically conditioned syncope". McCarthy's use of the term is distinct from Gouskova's (2003), which designates syncope that optimizes prosodic constraints.

Despite pervasive levelling, however, the language still retains some alternations. Most saliently, the restructured language has replaced the incipient rhythmic deletion pattern with deletion in the well-known two-sided open syllable (roughly VC_CV, Kuroda 1967), and inherited an apocope process (section 5.2.3). Concomitant with these changes, the language developed an innovative prefix system (Kaye 1974a, Piggott 1980:2, 1985b). Remarkably, despite pervasive levelling, there are paradigms in the modern language that require a composite UR, that is a UR that is not identical to any one surface form. This indicates that languages that undergo leveling are not bound to strictly concrete URs. Furthermore, the paradigms that require composite URs are exactly the ones in which apocope and two-sided open syllable syncope interact, which strongly suggests that URs and grammars are learned in tandem in both levelling stable scenarios.

As argued in chapter 4, if a learner cannot acquire the grammar that generates a language, then a plausible response is to acquire a grammar that generates only a portion of the language. Parallelist models of phonology cannot generate rhythmic syncope (Kager 1999, McCarthy 2008), as will be illustrated in section 5.6. If learners use a parallelist phonological architecture, they would have been forced to restructure Odawa. Attributing the cause of restructuring to the use of a parallelist grammar is circumstantially supported by the grammar that was acquired by the restructuring generation. Unlike rhythmic syncope, two-sided open syllable syncope and apocope can be represented with a parallelist grammar. In keeping with the notion that only a portion of an impossible language will be retained, this grammar only allows a subset of the alternations in a rhythmic syncope system to continue.

In order to keep the restructured and near-rhythmic syncope stages of Odawa distinct we will follow the practice of Richards (1997) and refer to the restructured language as "New Odawa" and the ancestral language as "Old Odawa". Examples are labeled with the stage of the language they belong to.

The chapter will procede as follows. Section 5.2 provides an overview of Old Odawa. Section 5.3 discusses the paradigm levelling and prefix reanalysis that mark New Odawa as a restructured language. Section 5.4 will describe the new two sided open syllable syncope process. Section 5.6 analyzes the shift to New Odawa using modern constraint based frameworks. Section 5.7 shows that the application of our learning approach to the Old Odawa data results in the New Odawa

grammar being a viable hypothesis.

5.1.1 Preliminaries

Because Old Odawa alternations are rather dramatic, mappings from underlying forms to surface forms for Old Odawa will be illustrated with rule-based derivations. The use of rules is not an endorsement of a rule-based theory, rather, they are employed simply as a convenient means of displaying complex alternations. Beyond the description of Old Odawa, constraint based approaches will be used, focusing mainly on classical Optimality Theory (OT, Prince and Smolensky 2004) and Harmonic Serialism (McCarthy 2008; 2010).

Some New Odawa phenomena in section 5.4.1 are variable. There are numerous proposals for generating variation in constraint-based phonology, including partial-order grammars (Anttila 1997), Stochastic OT (Boersma 1997, Boersma and Hayes 2001), Maximum Entropy Harmonic Grammar (Goldwater and Johnson 2003) and Noisy Harmonic Grammar (Boersma and Pater to appear). The data do not support a particular model of variation, but for continuity with the rest of the analysis, variable ranking as in Stochastic OT or partial-order grammars will be used.

5.1.1.1 Sources of Data

Prior descriptive scholarship on Odawa and other Ojibwe dialects is extensive. For brevity, we will only mention works that treat Odawa. Scholarship on Odawa and related Ojibwe dialects began in the 17th century (Hanzeli 1969) and was quite sophisticated by the 19th century, as demonstrated by Baraga's seminal grammar and dictionary (Baraga 1878b; 1878a). Baraga's work was followed by Bloomfield, whose work on Odawa from 1938 was posthumously published as Bloomfield (1957). See Goddard (1987) for a summary of Bloomfield's broad experience with Algonquian languages.

The latter half of the 20th century saw three major descriptive undertakings. First, the Odawa Language Project began carrying out fieldwork in 1968. This led to several publications, including Piggott et al. (1971), Piggott and Kaye (1973), Kaye (1974b) and Piggott (1980). Second, Rhodes began fieldwork in 1972, and ultimately authored a detailed description of Odawa morphosyntax (Rhodes 1976) and the authoritative *Eastern Ojibwa-Chippewa-Ottawa Dictionary* (Rhodes

1985a, henceforth the Rhodes dictionary). Finally, Valentine began fieldwork in 1983, producing a detailed dialectological survey (Valentine 1994), and a reference grammar (Valentine 2001).

The current study benefits from the consensus derived from such sustained investigation. Examples are drawn primarily from the Rhodes dictionary, which contains nearly 10,000 New Odawa entries and their corresponding Old Odawa forms. Additionally, examples were verified by consulting descriptions of non-syncopating dialects of Ojibwe. I consulted sources describing both the modern dialect spoken in Minnesota (Nichols and Nyholm (1995) or the *Ojibwe People's Dictionary* published by the University of Minnesota (2012) and older, geographically closer dialects (Baraga 1878b; 1878a). Also consulted were the analyses by Bloomfield (1957), Kaye (1973), Piggott (1980, 1983) and Valentine (2001). In the rare cases where a form cannot be found in the Rhodes dictionary, the consulted source is provided.

The data in the sources above were augmented by fieldwork carried out in the summer of 2011 at Walpole Island, Ontario. Three speakers of New Odawa who were born between 1938 and 1942 were interviewed. Data was collected on the productivity of two-sided open syllable syncope (section 5.4.1) and the distribution of person prefix allomorphs in New Odawa (section 5.3.2).

Because we are dealing with a case of language change, it is useful to provide some information on the sociological aspects of the speech community. Odawa is spoken in communities dispersed over several hundred miles of the Great Lakes region. Language loss confronts all communities to varying degrees. In some communities, like those located on Manitoulin Island, Ontario, intergenerational transmission was robust until approximately the 1970's. In other communities there have been no new native speakers since the 1940's. The speakers interviewed for this study belong to the last generation to speak Odawa natively on Walpole Island, Ontario. The learning situation was still fairly robust for these speakers, as some report being monolingual Odawa speakers until they began school.

The failure to acquire rhythmic syncope is unlikely to be related to non-linguistic factors like language attrition. Though it is true that Odawa communities have faced language attrition, restructuring has been observed even in robust communities. Indeed, as discussed in section 5.3.4, the first mention of restructuring is found in Piggott (1974 [1980]), which describes the speech of

one of the most robust communities, Manitoulin Island. Note further that even if learners did not have enough data to make a credible attempt at learning Old Odawa, the the restructurings of Old Russian and Old Irish (see section 5.5), which do not coincide with language attrition, must be explained.

Previous scholarship has shown Odawa to have highly complex derivational morphology. Much of the derivational morphology can be safely ignored, since nearly all etymologically multi-morphemic stems restructured without regard to the historical stem constituents. Additionally, many inflectional categories are unfamiliar to non-Algonquianists, so glosses will be fairly impressionistic.[2] Works discussing the morphosyntax of Algonquian languages include Valentine (2001), Ritter and Rosen (2005, 2009), Bruening (2009) and references therein. All relevant morpheme boundaries have been provided.

5.1.1.2 Phoneme Inventory and Syllable Structure

Both Old and New Odawa have a segment inventory of eighteen consonants and seven vowels. The IPA transcription used in this article follows the transcription used in the Rhodes dictionary and Valentine (2001), which represent a lenis-fortis contrast with the characters for voiced and voiceless segments, respectively.[3] For simplicity, we will often refer to lenis consonants as voiced, and fortis obstruents as voiceless, though nothing hinges on this terminology. The consonant inventory is shown in Figure (5.1).

The vowel system contrasts three short vowels and four long vowels, shown in Figure (5.2). The long vowels are more peripheral than the short vowels, and the low vowels are central. Nasalization

---

[2]Verbal affixes discussed in this paper include: *nɪ-* 'first person' (reanalyzed to *ndʌ-/ndoː-* in New Odawa), *gɪ-* 'second person' (reanalyzed to *gdʌ-/gdoː-* in New Odawa), *ʊ-* 'third person' (reanalyzed to null in New Odawa),-*d* (-*g* after nasals) 'third person conjunct order', *-aː* 'direct theme sign', *-n* 'singular inanimate object' *-wʌg* 'third person plural', *-ɪdɪzʊ* 'reflexive', *-wɪn* 'nominalizer', *-ʃkɪ* 'negative habit'. Nominal morphology includes: *nɪ-* 'first person' (reanalyzed to *ndʌ-/ndoː-* in New Odawa), *gɪ-* 'second person' (reanalyzed to *gdʌ-/gdoː-* in New Odawa), *ʊ-* 'third person' (reanalyzed to null in New Odawa), *-ʌg* 'animate plural' (*-wʌg* after an underlying short vowel), *-ʌn* 'inanimate plural' (*-wʌn* after an underlying short vowel). Acute accents mark stress, and underscores mark deleted segments.

[3]The only major departure in this transcription system from the traditional transcriptions used in Bloomfield (1957), Piggott (1980) and other early work is that fricative stop-sequences are represented with voiceless symbols instead of voiced symbols. Bloomfield indicates that fricative-stop clusters are phonetically intermediate between fortis and lenis. Such a designation probably comes from the fortis articulation of the fricative, but the unaspirated lenis-like articulation of the stop (Rhodes 1985a).

	Bilabial	Alveolar	Post-Alv.	Palatal	Velar	Glottal
Plosive	p b	t d			k g	ʔ
Nasal	m	n				
Affricate			t͡ʃ d͡ʒ			
Fricative		s z	ʃ ʒ			(h)
Glide	w			j		

Figure 5.1: Odawa Consonant Inventory

is contrastive on long vowels in stem-final position, though this is not included in Figure (5.2).
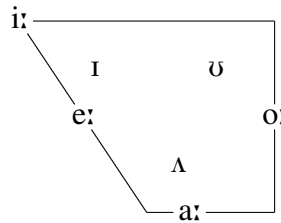
Figure 5.2: Odawa Oral Vowel Inventory

Before reduction became extreme in Old Odawa, the surface syllable structure consisted of vocalic nuclei and optional onsets and codas. Onsets maximally contained consonant-[w] sequences.[4] Word-internal syllables could only be closed by the first member of the sequences shown in Table (5.3). Word-final syllables could be closed by complex codas, consisting of the clusters shown in Table (5.3), though this was rare enough that some clusters are not attested word-finally.

Strident-stop	Nasal-obstruent
sp	mb
st	nd
sk	nz
ʃp	nʒ
ʃt	nd͡ʒ
ʃk	ng

Figure 5.3: Odawa Consonantal Clusters.

---

[4]There are two small wrinkles in the generalization. First, voiceless/fortis consonants were historically geminates and could not occur word-initially. A corollary of this is that word-initial branching onsets were only voiced stop-[w] sequences.

5.2 Old Odawa Description

Old Odawa was the language encountered by Bloomfield in 1938, a grammar of which was published as Bloomfield (1957). We will discuss four important processes in Old Odawa: the stress-reduction system, apocope, hiatus resolution and a rule of stem-initial [ʊ]-lengthening (Kaye 1973, Piggot 1980, Valentine 2001). The latter three processes were well-established parts of Odawa phonology, being attested in descriptions dating to the 1600s (Hanzeli 1969). Some of these processes can also be found in other Algonquian languages, suggesting a pre-historic origin. Drastic reduction was on the other hand a new phonetic change in progress, see section 5.2.1.1. For brevity and ease of identifying the effects of individual processes, Old Odawa will be displayed with rule based derivations. Section 5.6.1 will show how Old Odawa can be captured in modern constraint based systems.

5.2.1 Old Odawa Stress and Reduction

Old Odawa followed a typical iambic pattern, building right-headed feet from left to right (Hayes 1995). Stress was quantity sensitive, where only syllables containing long vowels were heavy. The final syllable of the word was always stressed, even at the cost of a degenerate foot. A gradient phonetic process severely reduced unstressed vowels to [ə], which we define as a centralized micro-vowel varying all the way down to zero in its duration. This is shown in (63). These reduced vowels were presumably highly likely to be misperceived as absent, which we represent with an additional line in derivations labeled as "likely percept". The gradience of the process in Old Odawa is underscored in section 5.2.1.1, and is given more formal consideration in section 5.6.1.

(63) Reduction

$$\begin{bmatrix} V \\ -\text{stress} \end{bmatrix} \rightarrow \quad ə$$

These points are illustrated in (64). The strong reduction process caused /mʌkɪzɪn-ʌn/ 'shoes' to surface as [mᵊkízᵊnʌ́n], which would be easily perceived as [mkɪznʌn].

(64) 'shoe-pl'   'shoe'   (O. Odawa)

/mʌkɪzɪn-ʌn/ /mʌkɪzɪn/   UR

(mʌkí)(zɪnʌ́n) (mʌkí)(zín) Stress

(mᵊkí)(zᵊnʌ́n) (mᵊkí)(zín) Reduction

[mᵊkízᵊnʌ́n]   [mᵊkízín]   SR

[mkɪznʌn]   [mkɪzɪn]   Likely percept

Closed syllables were not heavy and thus did not attract stress. Hence, as we see in (65), the initial syllable in /dʌngɪʃkʌmʌw-aː-d/ 'if he kicks his thing' had a coda but was still unstressed.

(65) 'If he kicks his thing'   (O. Odawa)

/dʌngɪʃkʌmʌw-aː-d/   UR

(dʌn.gíʃ)(kʌmʌ́)(wáːd) Stress

(dᵊn.gíʃ)(kᵊmʌ́)(wáːd) Reduction

[dᵊngíʃkᵊmʌ́wáːd]   SR

[dngɪʃkmʌwaːd]   Likely percept

The domain of stress assignment was the prosodic word, which included the stem, suffixes and person prefixes.[5] The person prefix inventory consisted of /nɪ-/ 'first person', /gɪ-/ 'second person', and /ʊ-/ 'third person'. Because person prefixes were monomoraic and footing applied from left to right, person prefixes shifted foot boundaries leftward, as shown in (66). Due to the interaction between stress and reduction, dramatically different sets of vowels surfaced in prefixed and unprefixed forms.

---

[5]There could only be one prefix included in the stress domain with a stem, since all other prefixes (called preverbs and prenouns in the Algonquianist literature) formed their own prosodic word. When these prefixes were employed, the person prefixes were attached to the preverb or prenoun instead of the stem, where they could be expected to trigger stress alternations within the prefix complex. Though we will not discuss preverbs or prenouns, their entries in the Rhodes dictionary indicate that they restructured in the same way that free morphemes did.

(66)   'I fish with a rod'   'If he fishes with a rod'   (O. Odawa)

/nɪ-gʊndʌmoːd͡ʒɪgeː/   /gʊndʌmoːd͡ʒɪgeː-d/   UR

(nɪgʊ́n)(dʌmóː)(d͡ʒɪgéː)   (gʊndʌ́)(móː)(d͡ʒɪgéːd)   Stress

(nᵊgʊ́n)(dᵊmóː)(d͡ʒᵊgéː)   (gᵊndʌ́)(móː)(d͡ʒᵊgéːd)   Reduction

[nᵊgʊ́ndᵊmóːd͡ʒᵊgéː]   [gᵊndʌ́móːd͡ʒᵊgéːd]   SR

[ngʊndmoːd͡ʒgeː]   [gndʌmoːd͡ʒgeːd]   Likely percept

The examples in (67) illustrate some of the variety of stem-internal alternations.

(67)   'I . . . '   'If he . . . '   (O. Odawa)

nᵊ-nísᵊdʊ́pw-áː   nᵊsídᵊpw-áː-d   'recognize his taste'

nᵊ-dʌ́gᵊnígéː   dᵊgʊ́nᵊgéː-d   'mix things'

nᵊ-bízᵊgéːʃín   bᵊzʊ́geːʃín-g   'stumble'

nᵊ-zʌ́nᵊgítóː   zᵊnʌ́gᵊtóː-d   'have a hard time'

nᵊ-gʊ́tᵊgʊ́mᵊnʌ́gᵊbín-áː   gᵊtígᵊmínᵊgíbᵊn-áː-d   'roll someone'

These alternations were quite widespread in Old Odawa. Unambiguous stem-internal alterna-
tions occurred in stems that began with a CV syllable. These constituted approximately 40% of
the lexicon.[6] More dramatic alternations were found in CVCV-initial words, which constituted
approximately 25% of the lexicon. The only words without alternations at the left edge began
with a heavy syllable. Though not the majority pattern, Old Odawa was rife with paradigmatic
alternations from the radical reduction process.

5.2.1.1 Severe Reduction as a Late Rule

Drastic reduction appeared fairly recently in Odawa. The first documentation of reduction comes
in a 1912 text collected by Sapir, though clearly an earlier date is possible. In this text some words
like *mɪw* and *nɪw* 'that other thing' appear in both syncopated and unsyncopated forms (Richard
Rhodes, p.c.). This indicates that reduction had yet to become fully pervasive. It is unlikely that

---

[6]These statistics are approximate, as there is only limited data on the Odawa lexicon at the time in question. These
figures were calculated from the *Ojibwe People's Dictionary* (University of Minnesota, 2012), which represents the
very closely related dialect spoken in Minnesota. There is no reason to expect a major difference between Odawa in
the 1930s and modern Minnesota Ojibwe in this regard.

reduction began much before Sapir's work, since Bloomfield's texts from a speaker born in 1868 show no radical reduction (Williams 1991), mirroring texts written by a native speaker in the late 19th century (Blackbird 1887). Furthermore, Baraga's dictionary admonishes readers that vowels in his transcription are "*never* silent" (Baraga 1878b:4, emphasis original), presumably to preclude pronunciations with "silent vowels" from English orthography, but underscoring a lack of drastic reduction.

Bloomfield encountered Old Odawa reduction in 1938. For his consultant, Old Odawa reduction was extreme and pervasive. Bloomfield described unstressed vowels as "rapidly spoken and often whispered or entirely omitted" (Bloomfield 1957:5). This observation is echoed by Kaye (1973) and Piggott (1980:81), in their description of data from speakers born in the early 20th century.

The gradual development of reduction into an ever more extreme process follows the typical trajectory of sound change. Bloomfield's consultant was at the extreme end of a phonetic continuum. In the final step of this change, children would misinterpret the severe but gradient reduction as categorical deletion. Whether children could represent and learn a rhythmic syncope system depends on whether the phonological component of UG allows serial evaluation. This will be spelled out more fully in section 5.6.

5.2.2 Old Odawa Hiatus Resolution

Old Odawa repaired vowel hiatus by epenthesizing [d] between the vowels.[7,8]

(68)   Hiatus Resolution

$\emptyset \rightarrow$ d / V _ V

---

[7]The epenthesis of [d] only resolved vowel hiatus that spanned the prefix-stem boundary. In other environments [ʔ] was epenthesized or one of the vowels deleted. These aspects of the phonology do not concern this discussion, making (68) an adequate representation for present purposes. Rule (60) in Piggott (1980) is the more precise articulation.

[8]An alternative analysis might posit that the first person prefix is underlyingly /nɪd-/, with deletion of the [d] preconsonantally. Such an analysis is proposed for Proto-Algonquian. However, I follow the analysis of underlying /nɪ-/ because of a distinct process of hiatus resolution on inalienably possessed nouns. In inalienably posessed nouns, the prefix vowel deletes to avoid hiatus. Hence, /nɪ-oːs/ 'my father', maps to [noːs]. Achieving the same surface form from underlying /nɪd-oːs/ is slightly more complicated. Nothing in the argument below hinges on this analysis.

Hiatus resolution was a well-established phonological process, and thus applied before phonetic reduction (69).

(69) 'My sacred story'   (O. Odawa)

/nɪ-aːdɪsoːkaːn/   UR

nɪ[d]aːdɪsoːkaːn   Hiatus Resolution

(nɪdáː)(dɪsóː)(káːn)   Stress

(nᵊdáː)(dᵊsóː)(káːn)   Reduction

[nᵊdáːdᵊsóːkáːn]   SR

[ndaːdsoːkaːn]   Likely percept

5.2.3 Old Odawa Apocope

Old Odawa had an apocope process that deleted word-final short vowels in inflected verbs.[9] Simplifying slightly, the rule is formalized in (70).

(70) Apocope

$V \rightarrow \emptyset / \_]_{V\ +person}$

Apocope is illustrated in (71), where the form on the left undergoes apocope while the form on the right does not because of the presence of a consonantal suffix.

(71) 'I am dry' 'If he is dry' (O. Odawa)

/nɪ-baːsʊ/ /baːsʊ-d/ UR

nɪbaːs_ — Apocope

(nɪbáːs) (báː)(sʊ́d) Stress

(nᵊbáːs) — Reduction

[nᵊbáːs] [báːsʊ́d] SR

[nbaːs] [baːsʊd] Likely percept

---

[9]Historically the process non-iteratively removed word-final short vowels and glides on all nouns and verbs (Piggott 1980). The rule appears to have been restructured some time before Old Odawa, since the only convincing alternations in the modern language are found in the deletion of vowels in inflected verbs.

The syllables moved into word-final position by apocope must have been stressed, because the mapping from underlying /nɪ-ɪnaːbeːwɪzɪ/ to [nᵊ[d]-ɪ́náːbéːwíz] 'I am shaped so' shows that they were not reduced. In a rule-based framework, this is handled by apocope applying prior to the stress rules. This is illustrated in (72).

(72) 'I am shaped so'   (O. Odawa)

nɪ-ɪnaːbeːwɪzɪ/   UR

nɪ[d]ɪnaːbeːwɪzɪ   Hiatus Resolution

nɪdɪnaːbeːwɪz_   Apocope

(nɪdɪ́)(náː)(béː)(wíz)   Stress

(nᵊdɪ́)(náː)(béː)(wíz)   Reduction

[nᵊdɪ́náːbéːwíz]   SR

[ndɪnaːbeːwɪz]   Likely percept

The examples in (73) demonstrate apocope in more forms.[10]

(73) 'If he ...'   'I ...'   (O. Odawa)

náːdáːsʊ́-d   nᵊ-náːdáːs   get things

néːwéːbí-d   nᵊ-néːwéːb   rest

níːgáːnᵊbízʊ́-d nᵊ-níːgáːnᵊbíz   drive ahead

áːbᵊdízí-d   n[d]-áːbᵊdíz   be useful

5.2.4 [ʊ] Lengthening

A rather unusual rule lengthened stem-initial [ʊ] after a prefix, which we formalize in (74).

(74) [ʊ] Lengthening

$$\text{ʊ} \rightarrow \text{oː} / \,]_{\text{prefix}} \_\_$$

Before reduction became extreme in Old Odawa this rule drove the alternation between *ʊpwaːgʌn* 'pipe' and *nɪ[d]-oːpwaːgʌn* 'my pipe', which is still observed in Minnesota Ojibwe. As reduction

---

[10]Our examples are restricted to showing [ʊ] and [ɪ] deleting because no verb stems or affixes ended in [ʌ]. Note however that the Ojibwe People's Dictionary does contain stems ending in [ʌ].

became more severe, Old Odawa surface forms came to have an alternation between [ə] and [oː], as (75) illustrates.

(75) 'pipe'   'my pipe'   (O. Odawa, Valentine 2001:62-63)

/ʊpwaːgʌn/   /nɪ-ʊpwaːgʌn/   UR

—   nɪoːpwaːgʌn   [ʊ] Lengthening

—   nɪ[d]oːpwaːgʌn   Hiatus Resolution

(ʊpwáː)(gʌ́n)   (nɪdóː)(pwáː)(gʌ́n) Stress

(əpwáː)(gʌ́n)   (nədóː)(pwáː)(gʌ́n) Reduction

[əpwáːgʌ́n]   [nədóːpwáːgʌ́n]   SR

[pwaːgʌn]   [ndoːpwaːgʌn]   Likely percept

5.3 Levelling and Recutting in New Odawa

New Odawa deviates from Old Odawa in two major respects. First, New Odawa has levelled out the Old Odawa alternations at the left edge of the word. Second, the person prefix system has undergone a radical change, with multiple innovative allomorphs for each prefix. Third, New Odawa possesses an innovative syncope process that is descended from Old Odawa reduction but is computed solely off of surface phonotactics, not stress. This section discusses the paradigm levelling and prefix re-analysis, while section 5.4 discusses the new grammar. In section 5.6, we argue that a parallelist architecture is sufficient to trigger restructuring and show how our learning algorithm correctly produces the observed levelling.

5.3.1 Loss of Stem Alternations

New Odawa has leveled out the Old Odawa alternations at the left edge of the stem. Recall that stems that began with a CV syllable in Old Odawa alternated dramatically. As seen in (76), when the word for 'arrive' was not prefixed, as in /dʌgʊʃɪn-g/ 'if he arrives', it surfaced as [dəgʊ́ʃín-g]. The addition of a prefix shifted the footing, resulting in [nə-dʌ́gəʃín] 'I arrive'.

(76)   'If he arrives'   'I arrive'   (O. Odawa)

      /dʌgʊʃɪn-g/   /nɪ-dʌgʊʃɪn/   UR

      (dʌgʊ́)(ʃíng)   (nɪdʌ́)(gʊʃín)   Stress

      (dᵊgʊ́)(ʃíng)   (nᵊdʌ́)(gᵊʃín)   Reduction

      [dᵊgʊ́ʃíng]   [nᵊdʌ́gᵊʃín]   SR

      [dgʊʃɪng]   [ndʌgʃɪn]   Likely percept

In New Odawa, prefixation no longer triggers the realization of a completely different set of vowels. Instead, the Old Odawa unprefixed form is used throughout the paradigm. Thus, the New Odawa word for 'if he arrives' is [dgʊʃɪn-g], while the main New Odawa correlate of Old Odawa [n-dʌgʃɪn] is [ndʌ-dgʊʃɪn]. The vowel-zero alternations of Old Odawa reduction have been lost in New Odawa. They have been replaced with a consistent vocalism derived mainly from Old Odawa unprefixed forms.

A second restructured feature of New Odawa is that the regular prefixation strategy has changed. Where Old Odawa had phonologically conditioned *n-* or *nd-*, New Odawa uses innovative prefixes *ndʌ-* or *ndoː-* interchangeably (Rhodes 1985a; 1985b, see also Valentine 2001), with some speakers also using *ndɪ-* as a less frequent variant.[11] Thus, in addition to [ndʌ-dgʊʃɪn] 'I arrive', [ndoː-dgʊʃɪn] and even [ndɪ-dgʊʃɪn] are grammatical. The origin and patterning of these innovative prefixes is discussed in section 5.3.2.

The loss of severe reduction alternations is exceptionless in New Odawa. For example, consider the Old Odawa words cited in (67), repeated in (77).

(77)   'I ...'   'If he ...'   (O. Odawa)

      nᵊ-nísᵊdʊ́pw-áː   nᵊsídᵊpw-áː-d   'recognize his taste'

      nᵊ-dʌ́gᵊnígéː   dᵊgʊ́nᵊgéː-d   'mix things'

      nᵊ-bíz²géːʃín   bᵊzʊ́geːʃín-g   'stumble'

      nᵊ-zʌ́nᵊgítóː   zᵊnʌ́gᵊtóː-d   'have a hard time'

      nᵊ-gʊ́tᵊgʊ́mᵊnʌ́gᵊbín-áː   gᵊtígᵊmínᵊgíbᵊn-áː-d   'roll someone'

---

[11]The Rhodes dictionary specifically lists *ndɪ-* for words that begin with the relative root *ɪʒ/ɪn*, instead of listing *nda-* which stands for any of *ndʌ-, ndɪ-, ndoː*. In my experience speakers do not indicate that *ndɪ-* is uniformly preferred with words beginning with the relative root.

Compare these forms with their productive New Odawa reflexes. In every case, the unprefixed form is the same as the Old Odawa unprefixed form, and this stem allomorph is used throughout the paradigm.

(78) 'I ...'   'I ...'   'If he ...'   (N. Odawa)

ndʌ-nsɪdpw-aː   ndoː-nsɪdpw-aː   nsɪdpw-aː-d   notice his taste

ndʌ-dgʊngeː   ndoː-dgʊngeː   dgʊngeː-d   mix things

ndʌ-bzʊgeːʃɪn   ndoː-bzʊgeːʃɪn   bzʊgeːʃɪn-g   stumble

ndʌ-znʌgtoː   ndoː-znʌgtoː   znʌgtoː-d   struggle

ndʌ-gtɪgmɪngɪbn-aː ndoː-gtɪgmɪngɪbn-aː gtɪgmɪngɪbn-aː-d   roll him

The loss of Old Odawa stem alternations in New Odawa is concomitant with the restructuring of prefixes, which we take up in the next section.

5.3.2 Prefix Recutting

The major New Odawa prefix allomorphs *ndʌ*-, *ndoː*-, and *ndɪ*- are unlikely to have been created *ex nihilo*, but rather arose through recutting, a diachronic shift in the placement of a morpheme boundary. A simple example of recutting is the shift in English from *a nadder* to *an adder* (Chantraine 1945; Lynch 2001; Diertani 2011).

Old Odawa stems that began with a short vowel followed by a heavy syllable (underlying /VCVV/) were particularly vulnerable to recutting in New Odawa, as their initial vowel only appeared when the stem was prefixed. Example (79) shows an Old Odawa derivation that motivated *ndʌ*-.

(79)   'I hang'   'If he hangs'   (O. Odawa)

/nɪ-ʌgoːd͡ʒɪn/   /ʌgoːd͡ʒɪn-g/   UR

nɪ[d]ʌgoːd͡ʒɪn   —   Hiatus Resolution

(nɪdʌ́)(góː)(d͡ʒín)   (ʌgóː)(d͡ʒíng)   Stress

(nᵊdʌ́)(góː)(d͡ʒín)   (ᵊgóː)(d͡ʒíng)   Reduction

[nᵊdʌ́góːd͡ʒín]   [ᵊgóːd͡ʒíng]   SR

[ndʌgoːd͡ʒɪn]   [goːd͡ʒɪng]   Likely percept

Because the stress algorithm "restarted" the iambic stress pattern after a long vowel, the person prefix was only able to affect the footing of the stem-initial vowel in /ʌgoːd͡ʒɪn/. This means that when the form had no person prefix, the stem-initial vowel was unstressed and severely reduced, producing *ᵊgóːd͡ʒíng* 'if he hangs'. When a person prefix was attached, the stem initial vowel was stressed, as in *nᵊdʌ́goːd͡ʒín*.

From the perspective of a language learner, the segmentation of *n-* given the forms *ndʌgoːd͡ʒɪn* 'I hang' and *goːd͡ʒɪng* 'if he hangs' leaves [dʌ] with no morphemic parse. In contrast, a historically incorrect parse of the prefix as *ndʌ-* leaves no unexplained material, as schematized in (80).

(80) | goːd͡ʒɪng   'If he hangs'
       ndʌ | goːd͡ʒɪn   'I hang'

With this segmentation, the formerly stem-initial vowel has become part of the prefix, and the stem is now consonant-initial.

This process can be repeated for stems that began with other short vowels. The derivations of the Old Odawa stem /ɪnaːbʌd͡ʒɪtoː/ 'use something so' in (81) demonstrate how *ndɪ-* arose. First, consider the derivation that generated the surface forms *ᵊn]'aːbᵊd͡ʒítóː-d* 'if he uses it so' and *nᵊ-dínáːbᵊd͡ʒítóː-n* 'I use it so'.

(81) 'I use it so'   'If he uses it so'   (O. Odawa)

/nɪ-ɪnaːbʌd͡ʒitoː-n/   /ɪnaːbʌd͡ʒitoː-d/   UR

nɪ[d]ɪnaːbʌd͡ʒitoːn   —   Hiatus Resolution

(nɪdí)(náː)(bʌd͡ʒí)(tóːn) (ɪmáː)(bʌd͡ʒí)(tóːd) Stress

(nˠdí)(náː)(bˠd͡ʒí)(tóːn) (ˠnáː)(bˠd͡ʒí)(tóːd) Reduction

[nˠdínáːbˠd͡ʒítóːn]   [ˠnáːbˠd͡ʒítóːd]   SR

[ndɪnaːbd͡ʒitoːn]   [naːbd͡ʒitoːd]   Likely percept

Aligning the shared material at the left edges of the words provides a prefix *ndɪ-* 'I'.

(82) ┃ naːbd͡ʒitoːd 'If he uses it so'

  ndɪ ┃ naːbd͡ʒitoːn 'I use it so'

Words like Old Odawa /ʊd͡ʒeːpɪzɪ/ 'be lively' created *ndoː-*. We first provide the Old Odawa derivations in (83). Recall from section 5.2.4 that [ʊ] lengthened after a prefix.

(83) 'I am lively'   'If he is lively'   (O. Odawa)

/nɪ-ʊd͡ʒeːpɪzɪ/   /ʊd͡ʒeːpɪzɪ-d/   UR

nɪoːd͡ʒeːpɪzɪ   —   [ʊ] Lengthening

nɪ[d]oːd͡ʒeːpɪzɪ   —   Hiatus Resolution

nɪdoːd͡ʒeːpɪz_   —   Apocope

(nɪdóː)(d͡ʒéː)(píz)   (ʊd͡ʒéː)(pɪzíd)   Stress

(nˠdóː)(d͡ʒéː)(píz)   (ˠd͡ʒéː)(pˠzíd)   Reduction

[nˠdóːd͡ʒéːpíz]   [ˠd͡ʒéːpˠzíd]   SR

[ndoːd͡ʒeːpɪz]   [d͡ʒeːpzɪd]   Likely percept

String alignment between the two surface forms favors a prefix *ndoː-* 'I' and a stem *d͡ʒeːpɪzɪ* 'be lively'.

(84) ┃ d͡ʒeːpzɪd 'If he is lively'

  ndoː ┃ d͡ʒeːpɪz 'I am lively'

108

Learners had evidence for simpler prefixes as well. The most frequent prefix allomorph *n-* was segmentable off of CVV-initial words like /gaːskʌnʊzʊ/ 'whisper', whose derivations appear in (85).

(85)   'I whisper'   'If he whispers'   (O. Odawa)

/nɪ-gaːskʌnʊzʊ/   /gaːskʌnʊzʊ-d/   UR

nɪ-gaːskʌnʊz‿   —   Apocope

(nɪgáːs)(kʌnʊ́z)   (gáːs)(kʌnʊ́)(zʊ́d)   Stress

(nˠgáːs)(kˠnʊ́z)   (gáːs)(kˠnʊ́)(zʊ́d)   Reduction

[nˠgáːskˠnʊ́z]   [gáːskˠnʊ́zʊ́d]   SR

[ngaːsknʊz]   [gaːsknʊzʊd]   Likely percept

String alignment of [ngaːsknʊz] 'I whisper' and [gaːsknʊzʊd] 'If he whispers' allows the segmentation of *n-*.

(86) | gaːsknʊzʊd   'If he whispers'
     n | gaːsknʊz   'I whisper'

Finally, the allomorph *nd-* would be pulled off of Old Odawa words that began with long vowels, like *nˠd-áːdáːgˠnéːʃín* 'I am snow-bound'.

(87)   'I am snow-bound'   'If he is snow-bound'   (O. Odawa)

/nɪ-aːdaːgʊneːʃɪn/   /aːdaːgʊneːʃɪn-g/   UR

nɪ[d]aːdaːgʊneːʃɪn   —   Hiatus Resolution

(nɪdáː)(dáː)(gʊné)(ʃín)   (áː)(dáː)(gʊné)(ʃíng)   Stress

(nˠdáː)(dáː)(gˠnéː)(ʃín)   (áː)(dáː)(gˠnéː)(ʃíng)   Reduction

[nˠdáːdáːgˠnéːʃín]   [áːdáːgˠnéːʃíng]   SR

[ndaːdaːgneːʃɪn]   [aːdaːgneːʃɪng]   Likely percept

The segmentation would be performed mechanically by string-alignment between [ndaːdaːgneːʃɪn] 'I am snow-bound' and [aːdaːgneːʃɪng] 'if he is snow-bound'.

(88) | aːdaːgneːʃɪng   'If he is snow-bound'
     nd | aːdaːgneːʃɪn   'I am snow-bound'

In sum, Old Odawa surface forms provided evidence for five prefix allomorphs for the first person; of these *ndʌ-* and *ndoː-* emerged as defaults.[12] Hence, in addition to the forms listed in (77), innovative prefixes appear on the words in (89), where we suppress *ndoː-* for brevity. Crucially, these innovative default prefixes appear on practically all stems. The data collected for this study indicate that default prefixes even occur on vowel-initial words, despite the fact that they always took *nd-* in Old Odawa.

(89)

If he ... (O/N. Odawa)	I ... (O. Odawa)	I ... (N. Odawa)	
gaːsknʊzʊ-d	n-gaːsknʊz	ndʌ-gaːsknʊz	whisper
naːbd͡ʒɪtoː-d	n[d]-ɪnaːbd͡ʒɪtoː-n	ndʌ-naːbd͡ʒɪtoː-n	use it so
aːbnʌmw-aː-d	n[d]-aːbnʌmw-aː	ndʌ-aːbnʌmw-aː	untie him
aːbdweːweːbzʊ-d	n[d]-aːbdweːweːbɪz	ndʌ-aːbdweːweːbɪz	make noise

The remaining prefix allomorphs are still in use to a limited degree (Rhodes 1985b, Bowers 2011, see also Kaye 1974a and Valentine 2001:62-72). The allomorph *nd-* sporadically appears on vowel-initial stems, bringing about words like *nd-aːbdweːweːbɪz* 'I make noise while moving'. Likewise, *n-* can appear on stems that begin with a vowel, a singleton consonant, or a legal branching onset, as seen in *n-aːbdɪz* 'I am useful', *n-koːd͡ʒiːʃɪw* 'I have lice', or *n-ʃkwaːtaː* 'I die'. Finally, the use of *ndɪ-* varies between speakers; some use it as a markedly less frequent alternative to *ndʌ-* and *ndoː-*, while others don't use it at all.

In contrast to Old Odawa, the default New Odawa prefix system attaches *ndʌ-* and *ndoː-* to all lexical items. This is a major restructuring of the prefix allomorph inventory. To deny that the language has radically changed its inventory would require proliferating word-initial vowels on all stems and arbitrarily forcing them to surface as [ʌ] or [oː], but not [ɪ] for the majority of speakers. Such an approach is not only suspect, it obtains only modest success, as historically hiatus avoiding allomorphs are predicted to become obligatory. This is disconfirmed by the attestation of *n-* as a secondary pattern that occurs even outside of its historical domain. A more accurate analysis states that prefixes in New Odawa have been recut, and they attach to a base that corresponds at the left edge to the Old Odawa unprefixed stem allomorph.

---

[12]Second person prefixes regularized in the same way, though third person inflection is often null, as the Old Odawa prefix *ʊ-* 'third person' left no segmental residue after syncope.

In sum, New Odawa restructuring features two concurrent phenomena. First, paradigms have been leveled, so that stem allomorphs that only appeared in the absence of prefixes now occur throughout the New Odawa paradigm. Second, the prefix inventory has been reshaped, gaining recut allomorphs.

5.3.3 New Odawa Prosody

Old Odawa reduction left a prosodic vacuum in New Odawa. By removing unstressed vowels, Old Odawa reduction thoroughly obscured prominence relations between syllables. Old Odawa was a system where escaping severe reduction was the most robust diagnostic for stress. As we have seen, New Odawa has lost the Old Odawa stress-conditioned alternations. Accordingly, New Odawa lacks evidence for Old Odawa stress. We may safely state that the Old Odawa iambic stress pattern has been lost.

Given the loss of unstressed vowels, the only remaining stress contrast in Old Odawa was between main stress and secondary stress. The main stress rule of Old Odawa targeted the antepenultimate foot (Kaye 1973, Piggott 1980, Halle and Vergnaud 1987). In Old Odawa surface forms, this translates straightforwardly into stress on the antepenultimate syllable. Valentine (2001) states that New Odawa also has antepenultimate stress.[13] My own field recordings confirm this. Thus, we find *áːbd͡ʒɪt͡ʃgʌn* 'tool', and *aːbd͡ʒít͡ʃgʌn-ʌn* 'tools'. Stress assignment interacts transparently with deletions that are still part of the modern language (discussed in section 5.4). This is exemplified by pairs like *déːweːgʌn* 'drum' and *déːweːg‿nʌn* 'drums'. Because stress assignment is completely transparent and plays no role in the remaining phenomena we will discuss, it will not be marked in the remaining discussion.

5.3.4 The Time Course of Restructuring

The approximate beginning of restructuring was identified by Piggott (1980:2), where affixation by speakers in their mid-thirties and younger on Manitoulin Island is noted to be considerably different

---

[13]The question may not be completely settled. Valentine (1994:156-7) expresses doubt that New Odawa has the Old Odawa main stress rule, though he does not specify what led to this conclusion.

from that of their elders. Kaye (1974a) and Rhodes (1985a,b) identify prefix restructuring as the crucial shift in affixation. Given that most of the fieldwork for Piggott (1980) was carried out in 1968-1970, the earliest that these New Odawa speakers could have been born is 1932.

Crucially, the early childhood of these speakers coincides with the severe reduction documented by Bloomfield's 1938 fieldwork. Recall from section 5.2.1.1 that the severity of this reduction had steadily increased from the early 1910's so that by 1938 outright deletion was common. These dates point to an important conclusion: the phonetic loss of underlying vowels and radical restructuring were temporally and causally linked.

The proposed sequence of events is as follows. The Old Odawa generation phonologically represented weak vowels, but produced them as voiceless, extra short, or even (as a free variant) entirely obscured segments. At this point, the language had arrived at the tipping point between gradient phonetic reduction and categorical phonological deletion. Children interpreted their primary language data as categorically syncopated. Concretely, Old Odawa speakers had phonological representations like [(mʌkí)(zín)] 'shoe' and [(nɪmʌ́)(kɪzín)] 'my shoe'. Children were exposed to drastic reduction, and represented the Old Odawa forms as categorically syncopated [mkɪzɪn] 'shoe' and [nmʌkzɪn] 'my shoe'. In response, children did not learn a system that reproduced the Old Odawa patterns.

5.4   New Odawa Grammar

This section investigates the phonological grammar of New Odawa. This rounds out the picture of New Odawa, showing that it is well-behaved despite the upheaval of the Old Odawa-New Odawa transition. In New Odawa, there are two processes that delete short vowels. Apocope deletes word-final short vowels, while syncope removes short vowels in the cross-linguistically common two sided open syllable environment (roughly, VC_CV, Kuroda 1967). Deletion in the two sided open syllable is traditionally understood as deletion being regulated by syllable structure constraints (Kisseberth 1970, Kenstowicz 1980, Gouskova 2003). Crucially, this does not require a serial stress-before-syncope grammar. This new syncope process is a categorical descendent from Old Odawa. My fieldwork revealed that even in very slow, monitored speech, deleted vowels are not

pronounced. In place of the stress-before-syncope grammar necessary to recapitulate Old Odawa, New Odawa developed a system of syncope insensitive to odd-even position, but instead regulated by surface phonotactics.

5.4.1 New Odawa Syncope Description

Syllabification is the major constraint on New Odawa short vowel deletion. Complex codas may be created by syncope, but must respect sonority sequencing constraints. Meanwhile, acceptable complex onsets include consonant-glide, strident-voiceless stop, and strident-stop-glide sequences. Syncope also cannot create word-final clusters. Syncope is obligatory at the right edge of the word, but is optional at the left edge.

This analysis uses URs that respect the alternation condition (Kiparsky 1968a; 1971). This takes the paradigm levelling observed in New Odawa at face value (see Albright 2002; 2005; 2010 for a potential mechanism that can force levelling to a single member of the paradigm). Vowels from Old Odawa prefixed forms that are absent from New Odawa paradigms are assumed to be absent from URs. Concretely, where Old Odawa had /dʌgʊʃɪn/ 'to arrive', New Odawa has /dgʊʃɪn/, because its paradigm contains *dgʊʃɪn-g* 'if he arrives' and *ndʌ-dgʊʃɪn.* A major effect of this is that underlying consonant clusters are common. Indeed, morpheme-internal VCVCV sequences are nearly nonexistent, so configurations for deletion are almost exclusively the result of morpheme concatenation.

In the simplest case, a vowel deletes between singleton consonants, as illustrated in (90).

(90)	Singular	Plural	(N. Odawa)
	aːnʌk	aːn.k-ʌg	brown thrasher
	waːgʊʃ	waːg.ʃ-ʌg	fox
	pwaːgʌn	pwaːg.n-ʌg	pipe
	nmeːgʊs	nmeːg.s-ʌg	brown trout
	tɪbdoːwʌn	tɪbdoː.w.n-ʌn	wheel
	mkɪzɪn	mkɪz.n-ʌn	shoe
	waːgkaːgʌn	waːgkaːg.n-ʌn	rainbow

113

Syncope also creates complex codas, as shown by the mapping from /aːbzɪŋgʊʃɪ-d/ to [aːbzɪŋg.ʃɪ-d] 'if he wakes up rested'. More examples are listed in (91).The deleted vowels in the underlying forms are attested in prefixed forms, which are included in (91). The apocope process that deletes the word-final vowel will be reviewed in section 5.4.2.

(91) | Underlying | Unprefixed | Prefixed | (N. Odawa) |
|---|---|---|---|
| aːndʌbɪ-d | aːnd.bɪ-d | ndʌ-aːndʌb | change seats |
| nɪʃpʌbɪ-d | nɪʃp.bɪ-d | ndʌ-nɪʃpʌb | sit so high |
| weːndɪzɪ-d | weːnd.zɪ-d | ndʌ-weːndɪz | be easy |
| nnɪŋgʌd͡ʒɪ-d | nnɪŋg.d͡ʒɪ-d | ndʌ-nnɪŋgʌd͡ʒ | shiver from cold |
| moːʃkʌmʊ-d | moːʃk.mʊ-d | ndʌ-moːʃkʌm | surface/appear |
| gwiːʃkʊʃɪ-d | gwiːʃk.ʃɪ-d | ndʌ-gwiːʃkʊʃ | whistle |

Syncope also creates complex onsets, as shown by the mapping from underlying /bʌgʌʃk-oːn/ 'cutweeds' to [bʌg.ʃk-oːn]. The Rhodes dictionary lists other words that form complex onsets when syncope applies, some of which appear in (92).

(92) | Singular | Plural | (N. Odawa) |
|---|---|---|
| nɪnɪ | nɪ.n-wʌg | man |
| waːwaːteːsɪ | waːwaːteː.s-wʌg | firefly |
| pɪt͡ʃɪ | pɪ.t͡ʃ-wʌg | robin |
| zaːdʌj | zaː.dj-ʌg | poplar |
| pʌkweːjʌʃk | pʌkweːj.ʃk-oːn | cattail |

Deletion is blocked when any other cluster would result. For instance, no alternation is observed between *d͡ʒiːgdʌbgʌn* 'broom' and *d͡ʒiːgdʌbgʌn-ʌn* 'brooms', where the vowel that does not delete is underlined. Many other words attest this blocking effect, among them *miːgaːdwɪn-ʌn* 'wars', *miːʒmɪn-ʌn* 'acorns', *bdʌknɪgʊ-d* 'if he has a nightmare' and *ʃɪdaːkpɪzʊ-d* 'if he is fastened'.

Syncope is also blocked when it would create otherwise legal nasal-voiced obstruent or strident-voiceless stop clusters at the right edge of the word. Hence, even though New Odawa has words

like *pʌkweːjʌʃk* 'cattail', or *nɪnwʌnʒ* 'milkweed', the vowels that separate [s‿k] in *ntaːdeːbweːsɪ̠k* 'if he is a good liar', or [n‿z] in *ndʌ-nd͡ʒɪnʌ̠z* 'I fight for a reason' surface faithfully.

Finally, in contrast to the obligatory syncope found at the right edge of the word, syncope is optional at the left edge of the word. In the fieldwork conducted for this study, New Odawa words that begin with CV syllables, like /ʒɪdaːbaːn/ 'to drag someone', freely vary when prefixed between alternants like [ndʌ-ʒɪdaːbaːn-aː] and [ndʌ-ʒdaːbaːn-aː] 'I drag him', or [ndʌ-nɪʃkweːm-aː] 'I bother him by talking' and [ndʌ-n.ʃkweːm-aː].[14]

5.4.2 New Odawa Apocope

Unlike syncope, which was significantly changed between New Odawa and Old Odawa, apocope was kept in the same form. That is, word-final short vowels delete in afffixed verbs. Hence, there are alternations like those shown in (93).

(93)	Non-Deleted	Deleted	(N. Odawa)
	bdʌknɪgʊ-d	ndʌ-bdʌknɪg	have a nightmare
	dgʊbɪ-d	ndʌ-dgʊb	sit with others
	gbeːʃɪ-d	ndʌ-gbeːʃ	seek shelter

Apocope differs from syncope in that apocope can create a word-final consonant cluster. Hence, there are alternations like *gɪniːwaːnzʊ-d* 'if he is rose colored' and *ndʌ-gɪniːwaːnz* 'I am rose colored'.

5.4.3 A Parallel OT Analysis

To model New Odawa syncope, a dispreference against short vowels must be overridden by phonotactic constraints enforcing cluster well-formedness and a ban on word-final clusters. We collapse the cluster well-formedness conditions into a single cover constraint LEGALMARGIN. Apocope requires that a constraint specifically banning word-final vowels outrank the ban on word-final

---

[14]The Old Odawa prefixed forms of these words were *nˀ[d]-ɪʒˀdaːbaːn-aː* and *nˀ[d]-oːnˀʃkweːm-aː*. This precludes the New Odawa syncopated forms being whole memorized Old Odawa forms.

clusters. To protect long vowels from deleting we assume that MAX-Vː, which prohibits the dele-
tion of long vowels, is high-ranking. (94) displays the crucial constraints for our discussion.

(94)   a. FINAL-C: Assign one violation mark for a prosodic word that does not end in a con-
          sonant (McCarthy 1993).

       b. *CC#: Assign one violation mark for a word-final consonant cluster.

       c. LEGALMARGIN (abbreviated as LEGMAR): Assign one violation mark for every com-
          plex coda that is not composed of glide-consonant, strident-voiceless stop ([sp, st, sk,
          ʃp, ʃt, ʃk]), or nasal-voiced obstruent ([mb, nd, nz, nʒ, nd͡ʒ, ng]), and every complex
          onset that is not composed of consonant-glide or [sp, st, sk, ʃp, ʃt, ʃk] (+ w, j).

       d. *V̆ (abbreviated as *V): Assign one violation mark for every short vowel.[15]

       e. MAX-V: assign one violation mark for every short vowel in the input that has no output
          correspondent.

5.4.3.1 Apocope Ranking

The rankings for the apocope alternation are straightforward, as FINAL-C must outrank MAX-V
for /ndʌ-bdʌknɪgʊ/ to map to [ndʌ-bdʌknɪg] 'I have a nightmare'.

(95)   FINAL-C ≫ *MAX-V (New Odawa)

ndʌ-bdʌknɪgʊ	FINAL-C	MAX-V
a. ☞ ndʌ-bdʌknɪg		*
b. ndʌ-bdʌknɪgʊ	*!	

---

[15]Gouskova (2003) argues that *V and other constraints from the *STRUC family should be excluded from CON.
Deletion under that theory is held to be driven by satisfaction of constraints that do not specifically penalize the
existence of structure. However, syncope sometimes only results in a complex onset (for instance, when CVwV maps
to C‿wV). In the absence of more evidence, such a deletion optimizes nothing other than a dispreference against short
vowels (though see Munshi and Crowhurst 2012). To be sure, not all CVwV strings syncopate, as discussed below, but
the central point is that *V or something like it is necessary. Whether this constraint is an ad-hoc creation of language
learners or part of universal CON is not a question we will pursue.

Furthermore, FINAL-C must also outrank *CC# for /ndʌ-gɪni:waːnzʊ/ to map to ndʌ-gɪni:waːnz]
'I am rose colored'.

(96)   FINAL-C ≫ *CC# (New Odawa)

ndʌ-gɪni:waːnzʊ	FINAL-C	*CC#	MAX-V
a. ☞ ndʌ-gɪni:waːnzʊ		*	*
b. ndʌ-gɪni:waːnzʊ	*!		

5.4.3.2 Syncope Ranking

The core ranking for the grammar is LEGALMARGIN, *CC# ≫ *V ≫ MAX-V. Because New
Odawa has underlying clusters that violate LEGALMARGIN, as seen in *dgʊʃɪn* 'he comes', or
waːbndaːng 'if he sees it', this constraint is not undominated. While these clusters are being sim-
plified by many speakers, this phenomenon is beyond the scope of this paper, and it is assumed
without further argument that MAX-C is high ranked.

The crucial ranking argument for LEGALMARGIN over *V in the grammar comes from word-
internal cluster avoidance. For instance, syncope cannot occur in d͡ʒiːgdʌbgʌn-ʌn because deletion
of the penultimate vowel in the word would create an illegal [bgn] cluster. We illustrate the block-
ing effect in tableau (97).

(97)   LEGALMARGIN ≫ *V (New Odawa)

d͡ʒiːgdʌbgʌn-ʌn	LEGMAR	*V	MAX-V
a. ☞ d͡ʒiːg.dʌb.gʌnʌn		***	
b. d͡ʒiːg.dʌb.gnʌn	*!	**	*

Because syncope does not create otherwise legal word-final clusters, *V must also be domi-

117

nated by *CC#. For instance, the vowel that separates [s‿k] in *ndʌ-ntaːdeːbweːsɪk* 'I am a good liar', surfaces faithfully, as shown by (98).

(98)   *CC# ≫ *V (New Odawa)

ndʌ-ntaːdeːbweːsɪk	*CC#	*V	MAX-V
a. ☞ ndʌntaːdeːbweːsɪk		*	
b. ndʌntaːdeːbweːsk	*!		*

*V must outrank MAX-V in the grammar, as seen in the mapping from /aːnʌk-ʌg/ 'brown thrashers' to [aːnkʌg] in (99).

(99)   *V ≫ MAX-V (New Odawa)

aːnʌk-ʌg	*V	MAX-V
a. ☞ aːn.kʌg	*	*
b. aː.nʌ.kʌg	**!	

Crucially, *V dominates other constraints that could regulate deletion, like *COMPLEXONSET or *COMPLEXCODA, because New Odawa deletion creates complex clusters so long as they are compliant with LEGMAR. This is shown by the mapping from /gwiːʃkʊʃɪ-d/ to [gwiːʃk.ʃɪ-d] 'if he whistles' in (100).

(100)   *V ≫ *ComplexCoda (New Odawa)

gwiːʃkʊʃɪ-d	*V	Max-V	*CompCoda
a. ☞ gwiːʃk.ʃɪd	*	*	*
b. gwiːʃkʊʃɪd	**!		

The tableau in (101) illustrates that *ComplexOnset must also be dominated by *V, as the mapping from /nɪnɪ-wʌg/ 'men' to [nɪnwʌg] shows that complex onsets can be formed by deletion.

(101)   *V ≫ *ComplexOnset (New Odawa)

nɪnɪ-wʌg	*V	Max-V	*CompOns
a. ☞ nɪ.nwʌg	*	*	*
b. nɪ.nɪ.wʌg	**!		

5.4.3.3   Optional Deletion at the Left Edge

During the fieldwork carried out for this study, two sided open syllable syncope was found to be optional at the left edge of the word. Since naturalistic production data is not currently available, it is difficult to give a complete account of the relative frequency of the variants. As an approximation, 21 out of 25 elicited stems with light initial syllables had an acceptable deletion variant (Bowers 2012§3.1.8).[16] Furthermore, deletion variants were judged to be as acceptable as a faithful variant roughly a quarter of the time.

To model this variation, we use variable ranking, as described by Anttila (1997) or work in

---

[16]Of the four stems that failed to show deletion, one of them, /gʊndaːgnaːpneː/ 'get a sore throat', would have had an unsyllabifiable intervocalic [gnd] sequence if deletion occured, as in *[ndʌ-gndaːgnaːpneː]. The failure of the other three stems to show deletion cannot be due to phonotactic constraints. We might explain this as optional deletion simply failing to apply.

Stochastic OT (Boersma 1997, Boersma and Hayes 2001), though Maximum Entropy (Goldwater and Johnson 2003) or Noisy Harmonic Grammar (Boersma and Pater to appear) could potentially also be used. Our analysis uses a positional faithfulness constraint MAX-V$_{\text{stem-initial}}$ (Beckman 1998), defined in (102).

(102)   MAX-V$_{\text{stem-initial}}$ (abbreviated as MAX-V$_S$): assign one violation mark if the first vowel in the stem in the input has no output correspondent.

Tableau (103) illustrates how the free-ranking of MAX-V$_S$ with *V enforces optional deletion between *ndʌ-kʌwaːteːʃɪn* and *ndʌ-kwaːteːʃɪn* 'I cast a shadow'. The jagged line represents free ranking.

(103)   *V variably ranked with MAX-V$_s$ (New Odawa)

ndʌ-kʌwaːteːʃɪn	MAX-V$_s$	*V	MAX-V
a. ☞ ndʌkʌwaːteːʃɪn		***(!)	
b. ☞ ndʌkwaːteːʃɪn	*(!)	**	*

The constraint rankings motivated for New Odawa are summarized in the Hasse diagram in (104); a dashed line indicates free ranking.

(104)   Hasse diagram of New Odawa rankings

FINAL-C

LEGMAR   *CC#MAX-V$_{\text{stem-initial}}$

*V

*COMPONS   MAX-V *COMPCODA
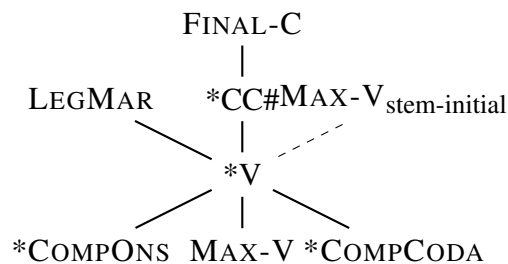
5.4.3.4 Non-Directionality of Syncope

The assumption that New Odawa syncope is regulated solely by phonotactics is confirmed by its non-directionality. Free variation in deletion sites is observed when more than one vowel is in a two-sided open syllable, licensing either one to delete. Such free variation is widespread in languages with phonotactic deletion, see the typological discussion in Bowers (2015) and the much-discussed case of French schwa deletion (Dell 1973, Kimper 2011 inter alia). Schematically, in the underlying configuration VCVCVCV either, but not both, of the underlined vowels may delete, producing VC_CVCV or VCVC_CV. This free variation in deletion sites could never have been produced in Old Odawa, because underlying strings were always footed in the same way.

In the fieldwork carried out for this study, free variation in deletion sites was observed. A consultant was asked to nominalize reflexive verbs, which created the underlying sequence /X-ɪdɪzʊ-wɪn/ 'X-reflexive-nominalizer'. This placed the second and third vowels of the reflexive morpheme -ɪdɪzʊ in the two-sided open syllable. On separate repetitions, the consultant produced forms where either vowel deleted, as shown in (105).[17] To ensure that the data fits on the page, we elide the root in the surface forms.[18]

(105)

Underlying	Surface 1	Surface 2	(N. Odawa)
daːngn-ɪdɪzʊ-wɪn	…-ɪd_zʊ-wɪn	…-ɪdɪz_-wɪn	self-feeling-ness
daːngd͡ʒiːbn-ɪdɪzʊ-wɪn	…-ɪd_zʊ-wɪn	…-ɪdɪz_-wɪn	self-brushing-ness

Other words were not recorded as having multiple variants, but nonetheless illustrate that deletion sites are not consistent when two vowels are in the two sided open syllable. Among them are [wiːkm-ɪd_zʊ-wɪn] 'self egging on-ness', [biːskʊnjeː-_dɪz_-wɪn] 'self clothing-ness' and [deːptaː-_d_zʊ-wɪn] 'self hearing from afar-ness'.

Formalizing the variation in deletion sites for our examples is not straightforward. The chief difficulty lies in the fact that intervocalic [d.zw] sequences are syllabifiable. Our account thus

---

[17]The underlying form /daːŋn/ differs from /daːngɪn/, which might be expected if the Rhodes dictionary is consulted. The speaker the verb was elicited from never gave any utterance that would have indicated that [ɪ] was present in the UR.

[18]Note that variants like *daːngníd_zʊwɪn* have deletion of the antepenultimate vowel, which would otherwise be stressed. Variants like *daːngnídɪz_wɪn* have deletion of a vowel that is two syllables away from the stressed vowel. If this analysis assumed that stress still conditioned deletion, these facts would be difficult to handle. See section 5.4.4.

predicts that *daːngn-ɪd‿z‿-wɪn, in which two vowels have been deleted, should be attested. We leave the precise formulation of this aspect of New Odawa grammar for further research. Note, however, that the larger point that New Odawa does not have the same grammar as Old Odawa stands, because the Old Odawa footing algorithm never produced variation in deletion sites.

5.4.3.5 Productive Extensions of Syncope

Rhodes (1985b) and Valentine (2001) state that in some cases, consonant cluster simplifications counterfeed syncope in New Odawa. However, for the consultants interviewed for this study, consonant cluster simplifications fed syncope. For instance, (106) shows that Old Odawa set up a [ds] cluster in all surface forms of /meːdʌsɪn/ 'miss him'.

(106) 'I miss him'   'If he misses him'   (O. Odawa)

/nɪ-meːdʌsɪn-aː/   /meːdʌsɪn-aː-d/   UR

(nɪméː)(dʌsí)(náː)   (méː)(dʌsí)(náːd)   Stress

(nᵊméː)(dᵊsí)(náː)   (méː)(dᵊsí)(náːd)   Reduction

[nᵊméːdᵊsínáː]   [méːdᵊsínáːd]   SR

[nmeːdsɪnaː]   [meːdsɪnaːd]   Likely percept

Rhodes reports that [d] frequently deletes before [s] in New Odawa, giving rise to the dictionary entry *meːsɪn-aː-d* 'if he misses him'.[19] The deletion of [d] places the [ɪ] in the two sided open syllable, but for this particular word, the dictionary indicates that the [ɪ] never deletes. The field data collected for this study show that in New Odawa, two-sided open syllable syncope has spread to this word. Hence, 'he misses him' was invariably pronounced *meːsn-aː*. Other cases where cluster simplification fed deletion are presented in (107).[20]

---

[19]In New Odawa the deletion of [d] before [s] is not universal, as the Rhodes dictionary also lists *mskʊdsʌn-g* 'if he paints it red' as an Odawa word.

[20]How widespread the new deletions are in New Odawa is not currently known. Importantly, the speakers who provided these data points are members of the first restructuring generation, and one of them contributed to the Rhodes dictionary. The discrepancy between the Rhodes dictionary and these data is most likely due to the regularization of exceptional forms. For additional discussion, see Bowers (2012)§3.1.9.

(107)	Old Odawa	New Odawa	Deletion Form	Gloss
	deːweːʔgʌn	deːweːgʌn	deːweːg‗n-ʌn	drum
	ʒɪbiːʔmʌw	ʒɪbiːmʌw	ʒɪbiːm‗w-aː	write for him

The extension of syncope to new contexts confirms that it is productive in New Odawa. Additional confirmation of the productivity of syncope comes from its application to nonce forms (Berko 1958). For instance, a speaker gave [deːtzɪs‗koːn] and [koːt͡ʃpaːk‗doːn] as the plurals for nonce [deːtzɪsɪk] and [koːt͡ʃpaːkʊd]. The plural forms lack a vowel that otherwise would be in a two sided open syllable, confirming that there is an active deletion process in the language.

5.4.3.6 Lexical Exceptions

Though syncope is very general in New Odawa, there are forms that do not straightforwardly conform. Most saliently, third person plural inflection does not condition syncope as would be expected given the analysis above. That is, third person plural verbs are formed with the suffix -*wʌg*, but stem-final short vowels frequently do not delete before the suffix. Thus, we see stem-final short vowels in the words in (108). These forms are from the author's field notes.

(108)	'They are white'	'They are alive'	(N. Odawa)
	/waːbaːnzʊ-wʌg/	/bmaːdɪzɪ-wʌg/	Underlying
	waː.baːn.zʊ-wʌg	bmaːd.zɪ-wʌg	Surface
	*waː.baːn.z-wʌg	*bmaːd.z-wʌg	Expected

The crucial aspect of (108) is that the stem-final vowels have not deleted even though the resulting string would be syllabifiable. The pattern seen in (108) is importantly not a direct holdover from Old Odawa, as illustrated in the derivations in (109). These Old Odawa forms are not present in Bloomfield (1957), they have instead been produced by applying rhythmic reduction to full voweled forms like those still found in Minnesota Ojibwe.

(109)   'They are white'   'They are alive'   (O. Odawa)

/waːbaːnzʊ-wʌg/   /bɪmaːdɪzɪ-wʌg/   UR

(wáː)(báːn)(zʊwʌ́g)   (bɪmáː)(dɪzí)(wʌ́g)   Stress

(wáː)(báːn)(zᵊwʌ́g)   (bᵊmáː)(dᵊzí)(wʌ́g)   Reduction

[wáːbáːnzᵊwʌ́g]   [bᵊmáːdᵊzíwʌ́g]   SR

[waːbaːnzwʌg]   [bmaːdzɪwʌg]   Likely percept

Comparison between (108) and (109), shows that in Old Odawa, the stem-final vowel in /waːbaːnzʊ-wʌg/ 'they are white' would have reduced, while in New Odawa it does not delete. This rules out the possibility that the persistence of stem-final vowels is entirely an inherited feature of Old Odawa.

Strikingly, some vowels do delete before -*wʌg*, as underlying /gaːweːʃkɪ-wʌg/ 'they are jealous' maps to [gaːweːʃk‿wʌg]. The situation is complicated further by deletion failing to occur in the closely parallel *gaːweːʃkɪ-wɪn* 'natural jealousy'. It is important to note that these exceptions are quite limited, apparently being confined to an unpredictable failure to delete before [w]-initial suffixes. These lexical idiosyncracies will need to be generated with an appropriate theory of lexical exceptions, see Pater (2010) for an approach that handles idiosyncratic syncope in Yine (Maipurean, Peru). It is possible as well that a study with a broader base of speakers will show these exceptions to be idiosyncratic to the speakers I have intereviewed.

This concludes the analysis of New Odawa. The essential point is that a novel syncope system has evolved out of the reduction patterns of Old Odawa. This system retains a subset of Old Odawa alternations (see section 5.7.2), but the modern pattern is governed by distinct principles.

5.4.4   No Rhythmic Syncope in New Odawa

Deletion patterns in New Odawa cannot be captured with iambic rhythmic syncope. This is because the modern system is insensitive to underlying even-odd position. Words like *ʒɪdaːbaːn-aː-d* 'if he drags him' and *deːweːg‿n-ʌn* 'drum-pl' illustrate this nicely. By virtue of following a long vowel, the deleted vowel in *deːweːg‿nʌn* is underlyingly in an "odd-numbered" position (as in *(déː)(wéː)(gʌnʌ́n)*). However, the first vowel in *ʒɪdaːbaːn-aː-d* is as well. Even more telling is the

124

free variation in deletion sites from section 5.4.3.4. Recall that /daːŋgn-ɪdɪzʊ-wɪn/ 'self-feeling-ness' varies between [daːŋgn-ɪd‿zʊ-wɪn] and [daːŋgn-ɪdɪz‿-wɪn]. In these examples, we see deletion of a vowel in either an even or an odd position. In contrast, phonotactic deletion describes the environment for deletion quite cleanly.

One might ask if deletion is regulated only with reference to the antepenultimate stress pattern of New Odawa. This does not work. Recall that stress falls on the surface antepenultimate syllable in New Odawa, as in *áːbd͡ʒɪt͡ʃgʌn* 'tool', and *aːbd͡ʒít͡ʃgʌn-ʌn* 'tool-pl'. Stress assignment interacts transparently with deletion, as seen in *déːweːgʌn* 'drum' and *déːweːg‿n-ʌn* 'drum-pl'. In other words, stress appears on the surface antepenultimate syllable even when deletion applies. Stress assignment accomodating deletion is *prima facie* evidence for stress following deletion in a serialist model. Of course, stress cannot condition syncope if it follows syncope. Deletion of ante-penultimate vowels further undermines this analysis. This can be seen in words like *ndʌ-ʒ‿biːw-aː* 'I draw him' (an optional variant of *ndʌ-ʒɪbiːwaː*) or *daːŋgn-ɪd‿zʊ-wɪn* 'self-feeling-ness'. If stress conditions deletion, the deletion of stressed vowels is highly anomalous.

5.5   Cross-Linguistic Responses to Rhythmic Syncope

Synchronic rhythmic syncope is strikingly absent from the languages of the world. Bowers (2015) discusses cases where the historical record shows that learners reject rhythmic syncope when it arises. We briefly consider the cases of Old Irish and Old Russian, which developed rhythmic syncope in approximately 550 CE and 1250 CE, respectively (Jackson 1953, McManus 1983, Kiparsky 1979:97-103).

5.5.1   Old Irish

Old Irish and other Insular Celtic languages had left-aligned trochees until unstressed vowels started dropping in 550 CE (Jackson 1953, McManus 1983). Though documents from the time do not discuss the phonetics of unstressed vowels, I assume here and below that this was extreme reduction like Old Odawa. Thurneysen (1946) provides the forms in (110).

(110)　　'Similar'　　'neg-similar-pl'

　　　　　　/kosamil/　　/e-kosamil-i/　　　UR

　　　　　　(kósa)(míl)　　(éko)(sámi)(lí)　　Stress

　　　　　　(kósᵊ)(míl)　　(ékᵊ)(sámᵊ)(lí)　　Reduction

　　　　　　[kosᵊmil]　　[ekᵊsamᵊli]　　　　SR

　　　　　　[kosmil]　　　[eksamli]　　　　　Likely percept

Thurneysen provides strong evidence that the language restructured. In many Old Irish manuscripts, deletion has ceased to be sensitive to the even-odd syllable count. In its place is deletion governed by phonotactics, as shown in (111).[21]

(111)　　Isolation　　Affixed

　　　　　kumaxtax　　kumax.t_x-u　　mighty, mightier

　　　　　komokʲus　　komoj.kʲ_s-e　　near, nearness

As can be seen above, if the resulting cluster was parsable into a simplex coda and an onset of rising sonority, deletion resulted.[22] The forms in (112) show deletion creating complex onsets of fricative-liquid and stop-fricative sequences, supporting a conjecture that indicate that in the restructured language complex onsets could maximally consist of stop-fricative-liquid sequences.[23]

(112)　　Historical　　　Restructured

　　　　　(tímᵊ)(θírext)　　tim.θ_rext　　　service

　　　　　(rágᵊ)(báθa)　　rag.b_θa　　　　have been sung

　　　　　(tárᵊ)(tíːˑetʲ)　　tar.t_satʲ　　　they have given

---

[21] One might expect [kumaxtax] 'mighty' and [komokʲus] 'near' to have consonant clusters in place of their second syllables. The [a] in [kumaxtax] survived because vowels never deleted before [xt] (Thurneysen 1946:67). The form *komokʲus*, originally had an additional syllable that was completely effaced, as some texts have orthographic {comḟocus}, where the diacritic on {ḟ} indicates that the segment has been deleted (see the Electronic Dictionary of the Irish Language, published by the Arts and Humanities Research Council 2013).The original parse was probably *(kóm_)(fókʲus)* 'near'.

[22] Deletion under-applies in *komokʲus* 'near', *soxumaxt* 'capable', and *foditiu* 'endurance'. Either these must be analyzed as exceptional forms, or a derived environment condition might have been active.

[23] The [e] of [tartiːˑet] becomes [a] in [tartsat] because [e] optionally dissimilates to [a] before a palatalized consonant (Thurneysen 1946:54).

126

The final evidence that Old Irish restructured comes from quadrisyllabic words where deletion of both medial vowels would result in an illegal cluster. Just as in New Odawa, free variation is observed here, even within the same manuscript. This is shown in (113).

(113)	Underlying	Variant 1	Variant 2
/tomonitis/	tom‗nitis	tomon‗tis	that they would think
/indirise/	ind‗rise	indir‗se	invade (participle)

The restructuring of Irish was abrupt. The first attestations of strong reduction date from the mid-sixth century, and by the mid-seventh century radical changes to Irish phonology had taken place (Jackson 1953, McManus 1983). If children resist positing rhythmic syncope, this rapid restructuring is expected. We need not follow Koch (1995:46) in stating that "it is simply impossible for a language to have evolved as quickly as the evidence seems to imply". To the contrary, very rapid change in response to rhythmic syncope is rather characteristic.

5.5.2 Old Russian

Old Russian had two short high vowels, [ɪ] and [ʊ], commonly referred to as *jers*. At the end of the Common Slavonic period, a process known as Havlik's Law strongly reduced jers in the weak branch of a trochaic foot and lowered them to [e, o] in the strong branch of a foot (Kiparsky 1979:97-103). Thus, Old Russian had alternations like those illustrated in (114).[24]

(114) | 'Hermit-acc.sg' | 'Hermit-nom.sg' | |
|---|---|---|
| /otʊʃʲɪlɪt͡sʲ-a/ | /otʊʃʲɪlɪt͡sʲ-ɪ/ | UR |
| (ótʊ)(ʃʲílɪ)(t͡sʲá) | (ó)(túʃʲɪ)(lít͡sʲɪ) | Stress |
| (ótʊ)(ʃʲélɪ)(t͡sʲá) | (ó)(tóʃʲɪ)(lét͡sʲɪ) | Lowering |
| (ótə)(ʃʲélə)(t͡sʲá) | (ó)(tóʃʲə)(lét͡sʲə) | Reduction |
| [otəʃʲeltət͡sʲa] | [otoʃʲəlet͡sʲə] | SR |
| [otʃʲelt͡sʲa] | [otoʃʲlet͡sʲ] | Likely percept |

It has been well established that Modern Russian and other Slavic languages no longer attest the historical pattern of jer deletion (Isačenko 1970, Kenstowicz and Kisseberth 1977, Pesetsky

---

[24]Kiparsky does not provide the actual phonetic forms of the words in (114), these are filled in from his description.

1979, Yearley 1995, Gouskova 2012). Hence, the alternation between *ot‿ʃʲelˆtsʲ-a* and *otoʃʲˍletsʲˍ* has been replaced with *otʃelʲˍtsˆ-a* and *otʃelʲetsˆ* in Modern Russian. Note that the stem-final vowel deletes in Modern Russian *otʃelʲˍtsˆ-a.* Importantly, Gouskova (2012) and Gouskova and Becker (2013) have shown that deletion is now regulated by phonotactic constraints that make no reference to stress.

Strikingly, Modern Russian "did not preserve *a single case* of multiple vowel/zero alternations" (Isačenko 1970:122, emphasis original). The loss of the Old Russian pattern was very swift, as Isačenko (1970:96) observes that "in most cases multiple vowel/zero alternations were eliminated simultaneously with the jer-shift itself". The one area where multiple vowel zero alternations endured was in prefixed forms, which Blumenfeld (2012) and Linzen et al. (2013) have shown to be highly lexicalized and subject to constraints that were not present historically.

5.5.3   Other Languages

The number of languages that have come to the brink of rhythmic syncope is prohibitively large to cover in depth. However, we note in passing that the following languages support the thesis that learners restructure rather than learn rhythmic syncope. Just within the Algonquian language family we find that Potawatomi, a very close relative of Odawa, developed severe reduction a short time before Odawa did. Hockett (1948:5) encountered restructuring in progress, as the unprefixed allomorphs of stems appeared after a recut prefix *ndoː-*, just as in New Odawa. The recut prefix has since become standard in the community, featuring even in language pedagogy (Robert Lewis, personal communication). Unami, an Eastern Algonquian language in the Delaware group, also levelled out most alternations that would evidence a prior rhythmic reduction stage (Goddard 1979:x-xxi).

Further afield, Malone (1997:143, 153) demonstrates that the variety of Mandaic spoken centuries ago had rhythmic syncope, but the modern language has syncope in the two-sided open syllable (see also Haberl 2009:96-98). Rhythmic syncope also appeared during Late Latin (Pope 1952, Rickard 1989, Jacobs 2004), but was lost by Gallo-Romance (Kager 1997, Jacobs 1989).[25]

---

[25]Archaic and Classical Latin had optional rhythmic syncope (Jacobs 2004, Blumenfeld 2006, Nishimura 2010; 2012). Its persistence can thus be explained within parallel OT with Output-Variant faithfulness (Kawahara 2002).

Some modern languages cited in support of synchronic rhythmic syncope by McCarthy (2008) can be shown to have syncope in the two sided open syllable instead. Ohala (1977) explicitly shows that schwa deletion in Hindi makes no reference to stress and occurs only in non-initial syllables when the resulting cluster is legal. Similarly, though descriptions of Tundra Nenets (Staroverov 2006, Kavitskaya and Staroverov 2008) make use of metrical structure, the alternations they discuss can also be the result of deletion being regulated by phonotactics. Neither Hindi nor Tundra Nenets attest the multiple vowel-zero alternations that are the hallmark of rhythmic syncope.

Other languages, like Aguaruna and Tonkawa, have been analyzed as having synchronic rhythmic syncope. In the case of Aguaruna, following Payne (1990), Alderete (2001), and McCarthy (2008), the language builds iambic feet from left-to-right. However, the language lacks prefixes (Overall 2007), so there is no morphological way for a change in foot assignment to ripple through the word, and all alternations are at the right edge of the word, much as in New Odawa. Furthermore, current dictionaries show an abundance of cases where syncope has either targeted an incorrect syllable or failed to target a particular syllable (Wipio Deicat 1996, Uwarai Yagkug and Paz Suikai 1998).[26] Finally, Tonkawa closely resembled rhythmic syncope, except a pattern of compensatory lengthening meant that deletion did not obscure the stress system (Hoijer 1933; 1946), that is, the deletion was not opaque. Consequently, the language did not show signs of restructuring. See Bowers (2015) for further discussion of these and similar cases.

5.6   Old Odawa in Constraints

The learning theory advanced in chapters 2 and 4 can produce New Odawa when exposed to paradigm-labeled data from what we might call Ostensible Old Odawa, that is, the likely percept of Old Odawa, which has categorical rhythmic syncope instead of vowel reduction. As a brief review, the learning theory starts with the basics of Classic OT: an enumeration of constraints which

---

[26]The same situation holds for Southeastern Tepehuan, viz. Willett (1982), Willett (1991), Kager (1999), and Summer Institute of Linguistics (2005). In Southeastern Tepehuan there is even the apparent deletion of long vowels, which makes the language anomalous amongst the vowel deletion systems discussed here. A clue to the solution might come from the fact that the apparent deletion is observed in reduplicated forms. This raises the possibility that the reduplication system could be re-analyzed along the lines proposed for the related Uto-Aztecan language Pima in Riggle (2006a).

must be ranked for parallel evaluation of linguistic forms. It then follows the procedure described in chapter 2 to incrementally accrue phonotactic rankings (ensuring that all surface forms are legal) and morphophonological rankings (ensuring that all faithfulness violations suggested by alternations are generated). In the event that no ranking can both ensure that the observed surface forms are legal and compel the faithfulness violations suggested by alternations, the learner shifts its focus away from generating all alternations to generating as many alternations as it can. Specifically, all rankings that are compatible with the phonotactic ranking requirements are enumerated and compared on the number of alternations that can be generated from the URs they permit from the privileged surface form of each paradigm. Our task is to show that no ranking of OT constraints generates the Old Odawa pattern of rhythmic syncope (Kager 1997, Jacobs 2004, see especially McCarthy 2008), and illustrate how the grammar shown in section 5.4 is a viable hypothesis under the relaxed requirements once the full language is determined to be impossible.

In this section we illustrate the above point in detail. First, we show the trivial result that Old Odawa as phonetic reduction is a possible system that adults could expose children to. However, if the reduced vowels are misperceived as deleted, a parallelist grammar is unable to capture the pattern, and restructuring must result.

5.6.1 Conjectured Adult Old Odawa

In this section, we describe Old Odawa as we presume it was represented by adults, focusing on the interaction between the stress pattern and reduction. As the reduction process was properly phonetic, we assign it to the phonetic module. Very few assumptions need to be made about the phonological component, other than that it can construct iambs from left-to-right.[27] The content of phonetics and phonology and their inter-relation is the subject of much debate (see Keating 1996, Flemming 2001 and Cohn 2007). I only assume that phonetics gradiently implements categorical phonological representations. For instance, an unstressed vowel can be realized with a duration anywhere between 0 milliseconds and whatever upper limit is necessary on language-specific grounds. With phonetic reduction and phonological stress, either serial or parallel evaluation can

---

[27]We are not concerned with how to force left-alignment. Under the theory proposed by Kager (2001), left-alignment is inevitable for iambic stress. Of course, alignment constraints can also be used.

be employed. For concreteness, I will use parallel evaluation.

To generate exhaustive parsing into iambs, IAMB and FOOTBINARITY must outrank TROCHEE, and EXHAUSTIVITY(word) must outrank FOOTBINARITY. These constraints are defined in (115).

(115) a. IAMB: Assign one violation mark for a foot that is not right-headed.

      b. TROCHEE: Assign one violation mark for a foot that is not left-headed.

      c. EXHAUSTIVITY(word): Assign one violation mark for a syllable that is not contained in a foot (Ito and Mester 2003, Selkirk 1995).

      d. FOOTBINARITY: Assign one violation mark for a non-branching foot.

The following tableau demonstrates the generation of the Old Odawa phonological representation [(mʌkí)(zín)] 'shoe', as shown in (116).

(116) EXHAUSTIVITY(word) ≫ FOOTBINARITY

      FOOTBINARITY, IAMB ≫ *TROCHEE (Old Odawa)

mʌkɪzɪn	EXHAUST	FTBIN	IAMB	TROCHEE
a. ☞ (mʌkí)(zín)		*		*
b. (mʌkí)zɪn	*!			*
c. (mʌ́)(kí)(zín)		**!*		
d. (mʌ́kɪ)(zín)		*	*!	

The phonological representation [(mʌkí)(zín)] must subsequently be realized phonetically. At this point, unstressed vowels may be extremely reduced. Abstracting away from the other facets of reduction, unstressed vowels will frequently be alloted no duration, and will never have a very long duration. Thus, Old Odawa speakers would have been observed saying [mkɪzɪn], while also producing [mᵊkɪzɪn]. Though this is only an approximation, we assume that the reduction was se-

vere enough to be misperceived as categorical deletion. Though parallel OT can handle reduction, as we will see in section 5.6, it cannot represent a rhythmic pattern of categorical deletion.

5.6.2 Parallelism and Ostensible Old Odawa

If extreme reduction were misperceived as categorical deletion, the ambient language might be best named as Ostensible Old Odawa. Crucially, to learn Ostensible Old Odawa, the deletion process would have to be placed in the phonological module with the stress process. Rhythmic syncope involves a rather unique interaction between stress and deletion. Stress feeds deletion, but deletion destroys the original foot patterns that it depends on. In other words, the stress pattern is negated on the surface, but is present in an intermediate representation. For this reason, serial evaluation is needed for rhythmic syncope. This is a point made cogently by McCarthy (2008). In other work (see Bowers 2015), I have argued that the failure to acquire rhythmic syncope shows that Harmonic Serialism makes false predictions concerning language change. Here, the only crucial point is that using parallel OT makes a learner be incapable of learning rhythmic syncope.

As a preview to section 5.7, we illustrate here the basic problem in generating Ostensible Old Odawa, a rhythmic syncope language rather than a reduction language, using parallelist phonology. Parallelist theories of phonology must assign foot structure and avoid unstressed vowels simultaneously. As McCarthy (2008:527) states, this means "classic OT cannot express the generalization that apparently underlies MCS [rhythmic syncope, DAB]". In an iambic language, rhythmic syncope is an alternation between surface full vowels in underlying "even" positions and surface null in underlying "odd" positions. The surface pattern of vocalism must be determined by the underlying count because the surface forms have had the count obscured by deletion. This point alone should raise alarm in an OT setting, since OT constraints are limited to faithfulness constraints that enforce similarity between inputs and outputs and markedness constraints that enforce surface phonotactic patterns. If the pattern of even-odd deletion were encoded in a single constraint, that constraint would specify a pattern of unfaithfulness between inputs and outputs. So-called two-level constraints substantially increase the power of the theory and are not standardly employed.

Using the standard constraints used by phonologists does not help either. The obvious way to

132

enforce binary counting is by first assigning feet, and syncopating from the weak branch of those feet. In a parallelist architecture, there is no prior guide to which vowels are unstressed, so deletion is not limited to a binary rhythmic pattern (Kager 1997, Jacobs 2004, Blumenfeld 2006, McCarthy 2008). To see this, consider the following constraints.

(117) a. WDCON: Words are parsed into prosodic words (and hence have at least one foot). Assign one violation mark for a word that contains no feet (Selkirk 1995).

  b. *V-Place$_{weak}$: Assign one violation mark for every full vowel in a weak metrical position (McCarthy 2008).

  c. MAX-V: Assign one violation mark for every vowel in the input that has no output correspondent.

  d. ID(stress): Assign one violation mark for every output vowel whose stress value does not match the stress value of its correspondent in the input.

In (118) we see that if we use the same constraints that succeeded for McCarthy's Harmonic Serialism analysis, the UR /mʌkɪzɪn/ does not map uniquely to [mkɪzɪn].

(118)   Failure of stress and deletion (Old Odawa phonologized)

mʌkɪzɪn	*V-PLACE$_{weak}$	ID(str)	FTBIN	MAX-V
a. ☞ (mkí)(zín)		**	**	**
b. (mʌkí)(zín)	*W	**e	*L	*L
c. ☞ (mʌ́k)(zín)	e	**e	**e	**e
d. (mʌ́)(kí)(zín)	e	***W	***W	L

The presence of an ERC with no W-assigning constraints signals that parallel OT fails no matter what constraint ranking is used. Intuitively, without intermediate representations to narrow down the candidate set and guide deletion, there is no way to select the correct vowel to delete.

Candidate (a) is the intended winner in (118) and avoids unstressed vowels, but there are other equivalently bad ways to avoid them, as candidate (c) shows.[28] The ultimate surface form is not decided by the constraints under consideration. Other constraints, like those on consonant clusters, would decide the winner. This is a particularly unwelcome result because candidate (c) has typologically less marked consonant clusters and could be expected to win. The problem is even more severe with longer words. Underlying /mʌzɪnʌʔɪɡʌn/ 'book' should surface as [(mzín)(ʔí)(ɡʌ́n)], but could just as easily surface as [(mzí)(nʌ́ʔ)(ɡʌ́n)], or [(mʌ́z)(nʌ́ʔ)(ɡʌ́n)].

5.7 Actuating Change

Thus far the chapter has taken pains to characterize the change that has been observed and establish the inability of parallel OT to generate the original Ostensible Old Odawa pattern. This section applies the theory of phonological learning developed in chapters 2 and 4 to Ostensible Old Odawa, and shows that the grammar defended in section 5.4 is rather highly valued. There are three key points to capture.

(119) a. Paradigm levelling occurred, with unprefixed forms as the pivot.

  i. Ostensible Old Odawa *mkɪzɪn, n-mʌkzɪn* '(my) shoe' > New Odawa *mkɪzɪn, ndʌ-mkɪzɪn*

 b. Two-sided open syllable syncope and apocope still drive alternations.

  i. New Odawa 'shoes' /mkɪzɪn-ʌn/ → [mkɪzn-ʌn]

  ii. New Odawa 'I am white' /ndʌ-waːbaːnzʊ/ → [ndʌ-waːbaːnz]

 c. New Odawa paradigms show evidence of composite URs.

  i. New Odawa 'be lively' /d͡ʒeːpɪzɪ/ *d͡ʒeːpzɪ-d, ndʌ-d͡ʒeːpɪz*

  ii. New Odawa 'play a game' /dnʌkmɪɡɪzɪ/ *dnʌkmɪɡzɪ-d, ndʌ-dnʌkmɪɡɪz*

---

[28]Changing the ranking to disallow degenerate feet does not help. Whether the winner of (118) is parsed as *(mkí)(zín)* or *(mkɪzín)*, there is no reason for the first underlying vowel to delete, as opposed to the second vowel with a parse like *(mʌkzín)*.

5.7.1 Prelude to Change: Detecting Inconsistency

Our learner must be supplied with constraints. We assume the inventory that was relevant for describing the New Odawa phonology in section 5.4, which is repeated below with the addition of DEP-V.

(120)   a. FINAL-C: Assign one violation mark for a prosodic word that does not end in a consonant (McCarthy 1993).

      b. *CC#: Assign one violation mark for a word-final consonant cluster.

      c. LEGALMARGIN (abbreviated as LEGMAR): Assign one violation mark for every complex coda that is not composed of glide-consonant, strident-voiceless stop ([sp, st, sk, ʃp, ʃt, ʃk]), or nasal-voiced obstruent ([mb, nd, nz, nʒ, nd͡ʒ, ng]), and every complex onset that is not composed of consonant-glide or [sp, st, sk, ʃp, ʃt, ʃk] (+ w, j).

      d. *V̆ (abbreviated as *V): Assign one violation mark for every short vowel.

      e. MAX-V: assign one violation mark for every short vowel in the input that has no output correspondent.

      f. DEP-V: assign one violation mark for every short vowel in the output that has no input correspondent.

Inconsistency results fairly quickly when this constraint set is deployed on Ostensible Old Odawa forms. First, we obtain phonotactic ERCs to ensure that observed surface forms are legal. (121) shows ERCs that are necessary for a word with a marked initial cluster and short vowels like *mkɪzɪn* 'shoe' to be legal.

(121)   Phonotactics From Ostensible Old Odawa

mkɪzɪn	Fɪɴᴀʟ-C	*CC#	Lᴇɢᴍᴀʀ	*V	Mᴀx-V	Dᴇᴘ-V
a. ☞ mkɪzɪn			*	**		
b. mɪkɪzɪn			L	***W		*W
c. mɪkzɪn			L	**	*W	*W
d. mkɪzn		*W	**W	*L	*W	

Candidate (b) in (121) shows that words like *mkɪzɪn* 'shoe' could have their word-initial cluster tolerated either because short vowels are too marked to break up the cluster or inserting them is forbidden. Encountering forms like *nʊkiː-d* 'if he works' provides evidence that short vowels are not too marked to break up clusters, as seen in (122).

(122)   Phonotactics From Ostensible Old Odawa

nʊkiːd	Fɪɴᴀʟ-C	*CC#	Lᴇɢᴍᴀʀ	*V	Mᴀx-V	Dᴇᴘ-V
a. ☞ nʊkiːd				*		
b. nkiːd			*W	L	*W	

Finally, word-final consonant clusters are attested in Ostensible Old Odawa surface forms like *ndoː-ɡniːwaːnz* 'I am rose colored', so *CC# may outrank Dᴇᴘ-V, as shown below.

136

(123)   Phonotactics From Ostensible Old Odawa

ndoːgniːwaːnz	Final-C	*CC#	LegMar	*V	Max-V	Dep-V
a. ☞ ndoːgniːwaːnz		*	*			
b. ndoːgniːwaːnzɪ	*W	L	*	*W		*W

The inability of the assumed constraint set to handle Ostensible Old Odawa comes to light once alternations are considered. For instance, once both *mkɪzɪn* 'shoe' and *nmʌkzɪn* 'my shoe' are observed, the learner needs to apportion unfaithfulness between the allomorphs while not contradicting the ranking requirements shown in (121-123). The inuitively obvious analysis is that when vowels of multiple different qualities alternate with zero, the grammar enforces deletion, and any observed vowel is underlying. In the case at hand, this means hypothesizing a UR /mʌkɪzɪn/ and deriving both the isolation allomorph *mkɪzɪn* and the post-prefixal allomorph *mʌkzɪn* from it. However, such a mapping is harmonically bounded under this constraint set, as shown in (124).

(124)   Attempted Morphophonology from Ostensible Old Odawa

mʌkɪzɪn	Final-C	*CC#	LegMar	*V	Max-V	Dep-V
a. ☹ mkɪzɪn			*	**	*	
b. mʌkzɪn			L	**	*	

Under this constraint set, it is possible to compel vowel deletion, but (124) shows that it is always better to delete a vowel if it will not violate constraints on consonant clusters. The intended output *mkɪzɪn* 'shoe' is a perpetual loser, and so this particular apportioning of unfaithfulness must

137

be abandoned. The remaining ways to characterize the alternations in the paradigm for 'shoe' all depend on epenthesis for one or both vowel-zero alternations, as reflected in the URs /mkzɪn/, /mʌkzɪn/ and /mkɪzɪn/. Each of these hypotheses can be made compatible with a few paradigms, but the ultimate fate for them must eventually be inconsistency. The reason for this is that epenthesis must introduce a vowel with a default quality, or a quality determined by the surrounding consonants. Since three different vowels alternate with zero in Ostensible Old Odawa, accounting for all alternations via epenthesis must fail.[29] Hence, while our mechanical learning process would have to churn through more data to arrive at the conclusion of global inconsistency, we can safely move on to exploring how the learner will act once the inconsistency of all hypotheses has been detected.

5.7.2 Effecting Change

Chapter 4 specifies that once inconsistency has been detected in all morphophonological hypotheses, the learner retracts morphophonological ERCs and enumerates the total rankings that are consistent with the phonotactic ERCs. To avoid visual clutter, the phonotactic ERCs collected in (121-123) were reduced to their Most Informative Basis (Brasoveanu and Prince 2011) and are presented in (125):

(125)	Final-C	*CC#	LegMar	*V	Max-V	Dep-V
			W	L	W	
			L	W		W
	W	L	L	W		W

It should be immediately clear that the constraint hierarchy summarized at the end of section 5.4 is compatible with the ERCs in (125). For convenience, we repeat (104) as (126), suppressing constraints that were not used during the phonotactic ranking discussion above, and making explicit the assumed high rank of Dep-V. The crucial point is that for every ERC in (125), the constraints

---

[29] An often overlooked possibility is that one alternating segment is epenthetic while other alternating segments are derived via deletion. For instance, /n-mkɪzɪn/ could map to [nmʌkzɪn] 'my shoe' via epenthesis of [ʌ] and deletion of [ɪ]. However, this analysis requires that [ɪ] and [ʊ] in stem-initial syllables be deleted, which, generalizing slightly from (124), is clearly a non-starter.

that assign L are dominated by (below) constraints that assign W in (126).

(126)   Hasse diagram of New Odawa rankings

DEP-V
|
FINAL-C
|
LEGMAR   *CC#
|
*V
|
MAX-V
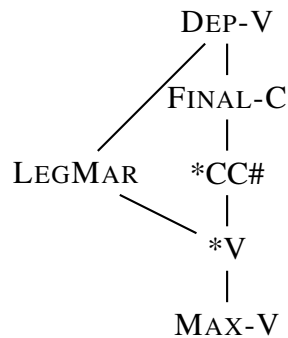
I currently lack simulation results to show that total rankings conforming to the dominance relations in (126) score better on the accuracy metric proposed at the end of chapter 4 (see algorithm 3) than other permutations that are compatible with the phonotactics rankings in (125). However, it should be clear that this grammar will score well, because while it does not enforce all of the deletions present in Ostensible Old Odawa, it does enforce deletion. Most importantly, the environments in which it enforces deletion correspond to deletion sites in Ostensible Old Odawa. For instance, take the New Odawa alternation between *mkɪzɪn* 'shoe' and *mkɪzn-ʌn* 'shoes'. The antecedents of these forms in Old Odawa are repeated from (64) below:

(127)   'shoe-pl'   'shoe'   (O. Odawa)

/mʌkɪzɪn-ʌn/   /mʌkɪzɪn/   UR

(mʌkí)(zɪmʌ́n)   (mʌkí)(zín)   Stress

(mᵊkí)(zᵊnʌ́n)   (mᵊkí)(zín)   Reduction

[mᵊkízᵊnʌ́n]   [mᵊkízín]   SR

[mkɪznʌn]   [mkɪzɪn]   Likely percept

The stem for 'shoe' /mʌkɪzɪn/ had odd parity underlyingly in Old Odawa. As a consequence of the stress rules, odd parity stems ended in a sequence of two stressed vowels, as seen in the parse *(mʌkí)(zín)*. When the plural suffix was appended to this word, the stem final syllable was parsed into the weak branch of a foot, though the penultimate syllable in the stem was still parsed into the strong branch of a foot. What was a result of the shifting stress value of stem-final syllables

conveniently matches the description of two-sided open syllable syncope, as shown by the mapping /mkɪzɪn-ʌn/ → [mkɪzn-ʌn].

Of course, Old Odawa even parity stems did not have allomorphy between singular and plural forms, and correspond to New Odawa stems where two-sided open syllable syncope is blocked by consonant cluster restrictions. For instance, the Old Odawa word for 'tool' was generated as follows:

(128) 'tool-pl'   'tool'   (O. Odawa)

/aːbʌd͡ʒɪt͡ʃɪɡʌn-ʌn/   /aːbʌd͡ʒɪt͡ʃɪɡʌn/   UR

(áː)(bʌd͡ʒí)(t͡ʃɪɡʌ́)(nʌ́n)   (áː)(bʌd͡ʒí)(t͡ʃɪɡʌ́n)   Stress

(áː)(bᵊd͡ʒí)(t͡ʃᵊɡʌ́)(nʌ́n) (áː)(bᵊd͡ʒí)(t͡ʃᵊɡʌ́n)   Reduction

[áːbᵊd͡ʒít͡ʃᵊɡʌ́nʌ́n]   [áːbᵊd͡ʒít͡ʃᵊɡʌ́n]   SR

[aːbd͡ʒɪt͡ʃɡʌn]   [aːbd͡ʒɪt͡ʃɡʌnʌn]   Likely percept

Even parity stems could not alternate in Old Odawa between singular and plural forms because no syllables changed their stress values between the singular and plural. Under the New Odawa analysis however, what had been even parity stems are now stems that have a consonant cluster before the stem-final syllable. As a result, the last short vowel in the stem cannot be placed in the two-sided open syllable syncope environment by the concatenation of a suffix, as confirmed by the New Odawa mapping /aːbd͡ʒɪt͡ʃɡʌn-ʌn/ → [aːbd͡ʒɪt͡ʃɡʌn-ʌn] 'tools'.

In short, by adopting a two-sided open syllable syncope grammar, a learner could account for at least some of the alternations that were present in Ostensible Old Odawa. The alternations between prefixed and unprefixed allomorphs are clearly beyond the capacity of such a grammar to generate, but if the presence or absence of a prefix is held constant, then alternations in suffixed forms can be generated. The only major question remaining is whether prefixed or unprefixed allomorphs should be prioritized. The levelling observed in New Odawa gives the answer away: unprefixed allomorphs were maintained over prefixed allomorphs. There are enough proposals that state that a morphologically primitive form should be chosen as the pivot in levelling that such a result is not surprising. However, note that Albright's theory that the phonologically most informative/least chaotic cell is chosen as the pivot may also predict this result, since preliminary calculations of

140

conditional entropy using a weighted non-deterministic finite state transducer indicate that the unprefixed → prefixed mapping is less chaotic than the prefixed → unprefixed mapping.

5.7.3 Composite URs in New Odawa

One of the more striking observations about New Odawa for the theory of levelling is that in the midst of a change whereby large portions of the lexicon shifted so that their URs became identical to their unprefixed allomorph, there is, as noted earlier, a class of words whose alternations in New Odawa strongly suggest a composite UR analysis. The words in question descend from Old Odawa even-parity verbs ending in two light syllables (and where the stem-final syllable was open). For instance, consider the Old Odawa paradigm for 'be smart', which had the following surface forms:

(129)   'If he is smart'   'I am smart'   (O. Odawa)

/ʊd͡ʒeːpɪzɪ-d/   /nɪ-ʊd͡ʒeːpɪzɪ/   UR

—   nɪoːd͡ʒeːpɪzɪ   [ʊ] Lengthening

—   nɪdoːd͡ʒeːpɪzɪ   Hiatus Resolution

—   nɪdoːd͡ʒeːpɪz‿   Apocope

(ʊd͡ʒéː)(pɪzíd)   (nɪdóː)(d͡ʒéː)(píz)   Stress

(ᵊd͡ʒéː)(pᵊzíd)   (nᵊdóː)(d͡ʒéː)(píz)   Reduction

[ᵊd͡ʒépᵊzíd]   [nᵊdóːd͡ʒéːpíz]   SR

[d͡ʒeːpzɪd]   [ndoːd͡ʒeːpɪz]   Likely Percept

The apocope process of Old Odawa removed the stem-final vowel in prefixed forms like *ndoːd͡ʒeːpɪz* 'I am smart', thereby making the penultimate syllable of these words become the last syllable in the word. Word-final syllables were stressed regardless of the even-odd count, so in prefixed forms the underlying penultimate vowel surfaced. In unprefixed forms however, the stem-final vowel was invariably protected from apocope, so the penultimate vowel was parsed into the weak branch of a foot and deleted, as can be seen in the form *d͡ʒeːpzɪ-d* 'if he is smart'. Crucially, whether the underlying penultimate vowel surfaced depended entirely on whether the underlying stem-final vowel had been apocopated. Hence, like many other paradigms in Ostensible Old Odawa, correctly generating all of the forms would require a composite UR.

Unlike most Ostensible Old Odawa paradigms that would require a composite UR, however, the New Odawa grammar can generate these alternations. In the unprefixed form, the penultimate vowel is in the two-sided open syllable, and deletes.

(130)   New Odawa Grammar Drives Syncope

d͡ʒeːpɪzɪ-d	Fɪɴ-C	*CC#	LᴇɢMᴀʀ	*V	Mᴀx-V
a. ☞ d͡ʒeːpzɪd				*	*
b. d͡ʒeːpɪzɪd				**!	

Meanwhile, apocope is also driven by the New Odawa grammar, which ensures that the Ostensible Old Odawa prefixed allomorph surfaces correctly.

(131)   New Odawa Grammar Drives Apocope

ndʌ-d͡ʒeːpɪzɪ	Fɪɴ-C	*CC#	LᴇɢMᴀʀ	*V	Mᴀx-V
a. ☞ ndʌd͡ʒeːpɪz			*	**	*
b. ndʌd͡ʒeːpɪzɪ	*!		*	***	
c. ndʌd͡ʒeːpzɪ	*!		*	**	*
d. ndʌd͡ʒeːpz		*(!)	**(!)	*	**

Note a point that has been implicit in all of the examples thus far. The UR for the stem, when placed in an appropriate context, is mapped to the appropriate Ostensible Old Odawa allomorph. This fact demonstrates that performing comprehension/recognition on the observed allomorph, as discussed by Eisner (2002) and Riggle (2004:194-198), produces a set which includes the UR as a possible underlying source for the allomorph. Because the UR is in the set of URs for both

the privileged allomorph and a non-privileged allomorph, the conditions for augmenting the accuracy measure discussed in chapter 4 are met here as well. The grammar's performance on these paradigms is another reason to expect that it would receive a high score under our evaluation measure.

5.7.3.1 A Minor Modification

Further examples from the class of words that require composite URs in New Odawa suggest that the method of calculating the accuracy of a grammar proposed in chapter 4 should be modified. The method as originally framed increases the measure of a grammar's accuracy when the set of URs for the privileged allomorph overlaps with the set of URs for a non-privileged allomorph (see section 4.7). However, words without a long vowel indicate that it may be better to score the amount of unexplained alternation when a UR for a privileged allomorph is mapped to a non-privileged allomorph.

An example that motivates such a proposal is the word for 'play a game', shown in (132)

(132) 'If he plays a game'   'I play a game'   (O. Odawa)

/dʌnʌkʌmɪgɪzɪ-d/   /nɪ-dʌnʌkʌmɪgɪzɪ/   UR

     —   nɪdʌnʌkʌmɪgɪz₋   Apocope

(dʌnʌ́)(kʌmɪ́)(gɪzɪ́d)   (nɪdʌ́)(nʌkʌ́)(mɪgɪ́z)   Stress

(dᵊnʌ́)(kᵊmɪ́)(gᵊzɪ́d)   (nᵊdʌ́)(nᵊkʌ́)(mᵊgɪ́z)   Reduction

[dᵊnʌ́kᵊmɪ́gᵊzɪ́d]   [nᵊdʌ́nᵊkʌ́mᵊgɪ́z]   SR

[dnʌkmɪgzɪd]   [ndʌnkʌmgɪz]   Likely Percept

In New Odawa, the forms of this word are *dnʌkmɪgzɪd* 'if he plays a game' and *ndʌ-dnʌkmɪgɪz* 'I play a game'. That is, New Odawa is evidently using the underlying form /dnʌkʌmɪgɪzɪ/. As (133 shows, the UR only drew the penultimate vowel from the Ostensible Old Odawa prefixed form.

(133)

	d		n	ʌ	k		m	ɪ	g	ɪ	z	ɪ		New Odawa UR
	d		n	ʌ	k		m	ɪ	g			z	ɪ	d	Ostensible O. Odawa SR
n	d	ʌ	n		k	ʌ	m		g	ɪ	z				Ostensible O. Odawa SR

More importantly, this UR could not have been in the set of URs for the Ostensible Old Odawa prefixed form under this grammar. This is made clear by the fact that the New Odawa prefixed allomorph generated by the grammar differs from the Ostensible New Odawa prefixed allomorph. This is easily seen by comparing the forms in (134).

(134)

ndʌ	d		n	ʌ	k		m	ɪ	g	ɪ	z		New Odawa SR
n		d	ʌ	n		k	ʌ	m		g	ɪ	z	Ostensible O. Odawa SR

If the solution that was evidently found by human learners is to be accurately reflected by our mechanistic learning process, it is necessary for individual URs to be scored for accuracy on the allomorphs that are derived from them under a grammar. The accuracy score of the grammar may then be determined from the scores of the URs. In keeping with the primacy of the privileged allomorph in our system and in language change, the URs to be scored should only be URs that the grammar maps to the privileged form.

Continuing with the example, there are at least four URs that the New Odawa grammar maps to the privileged allomorph *dnʌkmɪgzɪ-d*. These URs are shown in 135, where they are ordered by their relative similarity to *dnʌkmɪgzɪ-d* in the sense of similarity defined by Tesar (2013).

(135)
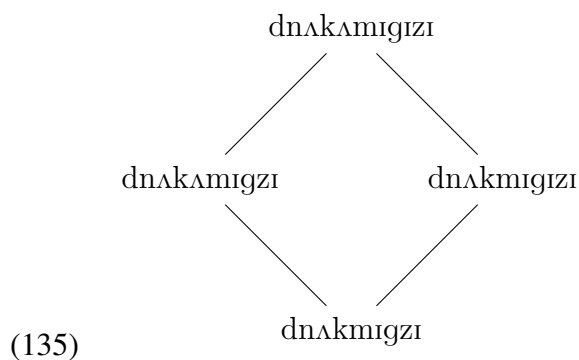
While these four URs just discussed all successfully map to the privileged surface form, they differ in what is produced for the prefixed allomorph. For instance, the URs /dnʌkmɪgzɪ/ and /dnʌkʌmɪgzɪ/ produce a large mismatch with the Old Odawa prefixed form:

(136)   New Odawa /ndʌ-dnʌkmɪgzɪ/ maps to [ndʌ-dnʌkmɪgz]

ndʌ-dnʌkmɪgzɪ	Fɪɴ-C	*CC#	LᴇɢMᴀʀ	*V	Mᴀx-V
a. ☞ ndʌdnʌkmɪgz		*	**	***	*
b. ndʌdnʌkmɪgzɪ	*!		*	****	

(137)   New Odawa /ndʌ-dnʌkʌmɪgzɪ/ maps to [ndʌ-dnʌkmɪgz]

ndʌ-dnʌkmɪgzɪ	Fɪɴ-C	*CC#	LᴇɢMᴀʀ	*V	Mᴀx-V
a. ☞ ndʌdnʌkmɪgz		*	*	***	**
b. ndʌdnʌkʌmɪgzɪ	*!		*	*****	

A closer match is obtained with the UR /dnʌkmɪgɪzɪ/, which contains the penultimate vowel:

(138)   New Odawa /ndʌ-dnʌkmɪgɪzɪ/ maps to [ndʌ-dnʌkmɪgɪz]

ndʌ-dnʌkmɪgɪzɪ	Fɪɴ-C	*CC#	LᴇɢMᴀʀ	*V	Mᴀx-V
a. ☞ ndʌdnʌkmɪgɪz			*	****	*
b. ndʌdnʌkmɪgɪzɪ	*!		*	*****	

Finally, with just the given constraints, the New Odawa grammar is unable to select a unique output for the UR /ndʌ-ndʌkʌmɪgɪzɪ/:

145

(139)   New Odawa /ndʌ-dnʌkʌmɪgɪzɪ/ maps to [ndʌ-dnʌkmɪgɪz] and [ndʌ-dnʌkʌmgɪz]

ndʌ-dnʌkmɪgɪzɪ	Fɪɴ-C	*CC#	LᴇɢMᴀʀ	*V	Mᴀx-V
a. ☞ ndʌdnʌkmɪgɪz			*	****	*
b. ☞ ndʌdnʌkʌmgɪz			*	****	*
c. ndʌdnʌkʌmɪgɪzɪ	*!		*	******	

It is not clear how to go about scoring an input that does not produce a unique output when considering the most immediately relevant constraints. With a larger constraint set, lower ranked constraints would eventually push the decision in different ways, depending on what total ranking was being scored. Within the scope of the current problem, it is important that the UR /dnʌkmɪgɪzɪ/ receive the best score, and not /dnʌkʌmɪgɪzɪ/. As shown in (140), however, the major variants generated from /dnʌkʌmɪgɪzɪ/ are on average closer to the Ostensible Old Odawa form. To avoid this UR edging out what appears to have been the choice ultimately taken by human learners of Odawa, I propose that the inaccuracies of the major variants derived from a UR are summed. In a larger constraint inventory where the ties are broken, it is unclear how to go about this, as the sums would need to be formed by adding the innacuracies produced by different rankings for the same UR.

(140)

n	d	ʌ	n		k	ʌ	m		g	ɪ	z	Ostensible O. Odawa	Inaccuracies
ndʌ	d		n	ʌ	k		m	ɪ	g		z	N. Odawa 1	5
ndʌ	d		n	ʌ	k		m	ɪ	g	ɪ	z	N. Odawa 2	4 ☞
ndʌ	d		n	ʌ	k		m	ɪ	g		z	N. Odawa 3	5
ndʌ	d		n	ʌ	k		m	ɪ	g	ɪ	z	N. Odawa 4	4 (7)
ndʌ	d		n	ʌ	k	ʌ	m		g	ɪ	z	N. Odawa 4	3 (7)

5.8 Local Summary

This concludes the discussion of Odawa. I have argued that as soon as learners perceived Old Odawa gradient reduction as categorical rhythmic syncope, massive restructuring resulted. A survey of similar cases indicates that this outcome is not unprecedented. This outcome may even be inevitable. The model advanced here explains the change as the best option for a learner that is confronted with an unlearnable system. Most importantly, the model is able to derive levelling while maintaining composite URs.

CHAPTER 6

Conclusion

This dissertation has developed a learner equipped with an explicit theory of synchronic grammar and a mechanism for recovering if it detects that its target language would require a grammar that contradicts its theory of synchronic grammar. The recovery mechanism is tailored to produce paradigm levelling, a type of historical change whose explanation has been sought since the Neogrammarians (see Kiparsky 1978). Case studies on Russian, Yiddish and Odawa indicate that our model is on the right track. Patterns of alternation that it predicts to be stable remain, whether it is for several centuries as in Russian, or through short periods of dramatic change as in Odawa. Meanwhile, patterns of alternation that it predicts to be unstable have been replaced, sometimes rapidly and dramatically, in the course of the history of a language. It would be imprudent to extrapolate from this sample that if a pattern is outside the generative capacity of mono-stratal Parallel OT supplied with markedness and input-output faithfulness constraints, then it will be re-analyzed. However, the discussed changes should also serve as a note of caution to the assumption that the apparent presence of a pattern that cannot be generated by OT is a sign of deficiency.

It is important to note that Stratal OT (Bermúdez-Otero 1999; 2006a, Kiparsky 2000) may derive some opaque phonology while still correctly predicting change in the cases above may. Opacity arises in Stratal OT through the serial interaction of phonology with morphology. Stem phonology applies in the morphological stem domain, word level processes apply when word-creating morphology is attached, and phrasal phonology applies in the phrasal domain. In order to preclude rhythmic syncope in Stratal OT, footing must not precede deletion. As it happens, in Old Odawa, Old Irish and Old Russian, footing and deletion both occurred in same domain. If learners can be prevented from opportunistically assigning processes with coextensive domains to different levels, then the rhythmic syncope cases discussed here will be excluded, as desired.

Linking serialism to morphology may provide a principled way to represent opacity while still excluding rhythmic syncope from the hypothesis space.

Restricting learners to parallelist grammars is a modest proposal. Phonological rhythmic syncope is a small part of the "derivational residue" in phonology (Hermans and van Oostendorp 1999). Much of the rest of the derivational residue is amenable to analysis with a wide range of proposals developed in the parallelist literature. These include distantal faithfulness (Kirchner 1995), *MAP constraints (Zuraw 2007), constraint conjunction (Moreton and Smolensky 2002), output-output correspondence (Benua 1997), turbidity (Goldrick and Smolensky 1999), allomorph listing (Sanders 2003), comparative markedness (McCarthy 2003), and targeted constraints (Wilson 2001).

The assumed theory of synchronic grammar and the accompanying learning model defended here are fairly crude. The constraints that define the parameters of linguistic variation are assumed to be given, and no attention has been given to the contents of these constraints other than to assume that they are finite state machines. At the very least, this work would profit from being integrated with work on constraint induction, see Heinz (2007; 2010), and Hayes and Wilson (2008). However, recent results show that relying on a set of constraints is not necessary to learn phonology. The work of Chandlee and colleagues (Chandlee 2014, Chandlee, Eyraud and Heinz 2014, Chandlee and Koirala 2014, Jardine, Chandlee, Eyraud and Heinz 2014) demonstrates how learning proceeds if phonology is limited to strictly local functions. Moreover, Cotterell, Peng and Eisner (2015) propose a learning algorithm that operates on paradigm-labeled data to discern underlying representations much as the model discussed here, but without relying on the constraints, or even the architecture of OT. Future work should apply these learning models to questions of historical change.

The proposed algorithm adopts the notion of a privileged cell from the Single Surface Base theory developed by Albright (2002; 2005; 2010). As it stands, this is an additional parameter that takes a significant amount of computation to determine. An obvious avenue for future work lies in assessing whether the privileged cell can be dispensed with entirely. In such a model, the final analysis will be a result only of the search for a maximally accurate, if still imperfect, grammar.

The most basic and important conclusion to be drawn comes from the case studies. Contrary to the predictions of previous work, paradigms that require composite URs are not especially unstable diachronically. The more important consideration is whether there exists a grammar that can compel the observed paradigmatic alternations. When there is no such grammar, the UR changes and levelling results. Parallel results from other researchers suggests that this view may be an emerging consensus (see Bermúdez-Otero in prep; 2014b).

BIBLIOGRAPHY

Ackerman, F., J. Blevins, and R. Malouf (2009). Parts and wholes: Implicative patterns in inflectional paradigms. In J. Blevins and J. Blevins (Eds.), *Analogy in Grammar: Form and Acquisition*, pp. 54–82. Oxford University Press.

Albright, A. (2002). *The Identification of Bases in Morphological Paradigms*. Ph. D. thesis, University of California, Los Angeles.

Albright, A. (2004). Sub-optimal paradigms in Yiddish. In V. Chand, A. Kelleher, A. Rodríguez, and B. Schmeiser (Eds.), *WCCFL 23*, pp. 1–14. Cascadilla Press.

Albright, A. (2005). The morphological basis of paradigm leveling. In L. Downing, T. Hall, and R. Raffelsiefen (Eds.), *Paradigms in Phonological Theory*. Oxford University Press.

Albright, A. (2008). Inflectional paradigms have bases too: Evidence from Yiddish. In A. Bachrach and A. Nevins (Eds.), *The Bases of Inflectional Identity*. Oxford University Press.

Albright, A. (2010). Base-driven leveling in Yiddish verb paradigms. *Natural Language and Linguistic Theory 28*, 475–537.

Albright, A. and B. Hayes (2002). Modeling English past tense intuitions with minimal generalization. In *Proceedings of the 2002 Workshop on Morphological Learning*. Association of Computational Linguistics.

Albright, A. and B. Hayes (2003). Rules vs analogy in English past tenses: A computational/experimental study. *Cognition 90*, 119–161.

Albright, A. and Y. Kang (2008). Predicting innovative variants in Korean verb paradigms. In *Proceedings of CIL18: The 18th International Congress of Linguists*.

Alderete, J. (2001). *Morphologically Governed Accent in Optimality Theory*. Outstanding Dissertations in Linguistics. Routledge.

Angluin, D. (1980). Inductive inference of formal languages from positive data. *Information and Control 45*, 117–135.

Anttila, A. (1997). Deriving variation from grammar: A study of Finnish genitives. In F. Hinskens, R. van Hout, and L. Wetzels (Eds.), *Variation, Change and Phonological Theory*. Amsterdam: John Benjamins.

Arts and Humanities Research Council (2013). Electronic dictionary of the irish language.

Avanesov, R. I. (1956). *Fonetika Sovremennogo Russkogo Literaturnogo Iazyka*. Izdatel'stvo Moskovskogo Universiteta.

Avanesov, R. I. (1972). *Russkoe Literaturnoe Proiznoshenie*. Prosveshchenie.

Avanesov, R. I. (1985). Information on pronunciation and stress . In *Orthoepical dictionary of the Russian language. Pronunciation. Stress. Grammatical forms* . . . . Russian Language. In Russian.

Bailey, R. W. (1996). *Nineteenth-Century English*. University of Michigan Press.

Bakovic, E. (2007). A revised typology of opaque generalizations. *Phonology 24*(2), 217–259.

Bakovic, E. (2011). Opacity and ordering. In J. Goldsmith, J. Riggle, and A. Yu (Eds.), *The Handbook of Phonological Theory* (Second ed.). Wiley-Blackwell.

Baraga, F. (1850 [1878]b). *A Theoretical and Practical Grammar of the Otchipwe Language* (Second ed.). Beauchemin and Valois.

Baraga, F. (1853 [1878]a). *A Dictionary of the Otchipwe Language: Explained in English* (Second ed.). Beauchemin and Valois.

Baudouin de Courtenay, J. (1895 [1972]). Versuch einer Theorie phonetischer Alternationen. In E. Stankiewicz (Ed.), *A Baudoin de Courtenay Anthology*, pp. 144–212. Indiana University Press.

Beckman, J. (1998). *Positional Faithfulness*. Ph. D. thesis, University of Massachusetts.

Benua, L. (1997). *Transderivational Identity: Phonological Relations between Words*. Ph. D. thesis, University of Massachusetts, Amherst.

152

Berko, J. (1958). The child's learning of English morphology. *Word 14*, 150–177.

Bermúdez-Otero, R. (1999). *Constraint Interaction in Language Change: Quantity in English and Germanic*. Ph. D. thesis, University of Manchester.

Bermúdez-Otero, R. (2006a). Morphological structure and phonological domains in Spanish denominal derivation. In S. Colina and F. Martínez-Gil (Eds.), *Optimality-Theoretic Studies in Spanish Phonology*. John Benjamins.

Bermúdez-Otero, R. (2006b). Phonological change in optimality theory. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (Second ed.), Volume 9, pp. 497–505. Elsevier.

Bermúdez-Otero, R. (2007). Diachronic phonology. In P. de Lacy (Ed.), *The Cambridge Handbook of Phonology*. Cambridge University Press.

Bermúdez-Otero, R. (2011). Cyclicity. In M. van Oostendorp, C. Ewen, E. Hume, and K. Rice (Eds.), *The Blackwell Companion to Phonology*, Volume 4, Chapter 85. Blackwell.

Bermúdez-Otero, R. (2014a). Amphichronic explanation and the life cycle of phonological processes. In P. Honeybone and J. C. Salmons (Eds.), *The Oxford handbook of Historical Phonology*. Oxford University Press.

Bermúdez-Otero, R. (2014b). French adjectival liaison: Evidence for underlying representations. Handout Given at Oxford University.

Bermúdez-Otero, R. (In prep). Stratal phonology: Arguments for cyclic containment, morphological implications. University of Manchester ms.

Bermúdez-Otero, R. and R. M. Hogg (2003). The actuation problem in Optimality Theory: Phonologization, rule inversion and rule loss. In D. E. Holt (Ed.), *Optimality Theory and Language Change*, pp. 91–119. Kluwer Academic Publishers.

Blackbird, A. J. (1887). *History of the Ottawa and Chippewa Indians of Michigan; A Grammar of their Language and Personal and Family History of the Author*. Ypsilantian Job Printing House.

Blevins, J. (2004). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge University Press.

Bloomfield, L. (1957). *Eastern Ojibwa: Grammatical Sketch, Texts and Word List*. Ann Arbor: University of Michigan Press.

Blum, L. and M. Blum (1975). Toward a mathematical theory of inductive inference. *Information and Control 45*, 125–155.

Blumenfeld, L. (2006). *Constraints on Phonological Interactions*. Ph. D. thesis, Stanford.

Blumenfeld, L. (2012). Vowel-zero alternations in Russian prepositions: Prosodic constituency and productivity. In V. Makarova (Ed.), *Russian Language Studies in North America: New Perspectives from Theoretical and Applied Linguistics*, pp. 43–69. Anthem Press.

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

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

Boersma, P. and J. Pater (To Appear). Convergence properties of a gradual learning algorithm for Harmonic Grammar. In J. McCarthy and J. Pater (Eds.), *Harmonic Grammar and Harmonic Serialism*. Equinox Press.

Bowers, D. (2011, September). Odawa field notes. UCLA manuscript.

Bowers, D. (2012). Phonological restructuring in Odawa. Master's thesis, University of California, Los Angeles.

Bowers, D. (2015). Phonological restructuring in Odawa. UCLA ms.

Brasoveanu, A. and A. Prince (2011). Ranking and necessity: The Fusional Reduction algorithm. *Natural Language and Linguistic Theory 29*, 3–70.

Bruening, B. (2009, March). Algonquian languages have A-movement and A-agreement. *Linguistic Inquiry 40*, 427–445.

Burzio, L. (1998). Multiple correspondence. *Lingua 104*, 79–109.

Burzio, L. (2002). Surface-to-surface morphology: When your representations turn into constraints. In P. Boucher (Ed.), *Many Morphologies*, pp. 142–177. Cascadilla Press.

Bybee, J. (1985). *Morphology: A Study of the Relation between Meaning and Form.* John Benjamins.

Chandlee, J. (2014). *Strictly Local Phonological Processes.* Ph. D. thesis, University of Delaware.

Chandlee, J., R. Eyraud, and J. Heinz (2014). Learning strictly local subsequential functions. In *Transactions of the Association for Computational Linguistics*, Volume 2, pp. 491–503.

Chandlee, J. and C. Koirala (2014). Learning local phonological processes. In *University of Pennsylvania Working Papers in Linguistics*, Volume 20.

Chantraine, P. (1945). *Morphologie Historique du Grec.* Librairie Klincksieck.

Chomsky, N. and M. Halle (1965). Some controversial questions in phonological theory. *Journal of Linguistics 1*, 97–138.

Chomsky, N. and M. Halle (1968). *The Sound Pattern of English.* Harper and Row.

Cohn, A. (2007). Phonetics in phonology and phonology in phonetics. In *Working Papers of the Cornell Phonetics Laboratory*, Volume 16, pp. 1–31. Cornell.

Corbiere, M. A. Introduction to Ojibwe. Ms. University of Sudbury.

Cotterell, R., N. Peng, and J. Eisner (2014). Stochastic contextual edit distance and probabilistic FSTs. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Volume 2, ACL'14*, pp. 625–630.

Cotterell, R., N. Peng, and J. Eisner (2015). Modeling word forms using latent underlying morphs and phonology. In *Transactions of the Association for Computational Linguistics*.

Crosswhite, K. (1999). *Vowel Reduction in Optimality Theory*. Ph. D. thesis, University of California, Los Angeles.

Daland, R., B. Hayes, J. White, and M. Garellek (2011). Explaining sonority projection effects. *Phonology 28*, 197–234.

De Chene, B. (2010, May). Description and explanation in inflectional morphology: The case of the Japanese verb. Waseda University manuscript.

de Saussure, F. (1916 [1959]). *Course in General Linguistics*. The Philosophical Library.

Dell, F. (1973). *Les règles et les sons: Introduction à la phonologie générative*. Paris: Hermann.

Diertani, C. E. A. (2011). *Morpheme Boundaries and Structural Change: Affixes Running Amok*. Ph. D. thesis, University of Pennsylvania.

Eisner, J. (2002). Comprehension and compilation in Optimality Theory. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 56–63.

Flemming, E. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology 18*, 7–44.

Fruehwald, J. (2013). *The Phonological Influence on Phonetic Change*. Ph. D. thesis, University of Pennsylvania.

Garrett, A. (2008). Paradigmatic uniformity and markedness. In J. Good (Ed.), *Explaining Linguistic Universals: Historical Convergence and Universal Grammar*, pp. 125–143. Oxford University Press.

Gess, R. (2003). On re-ranking and explanatory adequacy in a constraint-based theory of phonological change. In *Optimality Theory and Language Change*. Kluwer Academic Publishers.

Goddard, I. (1979). *Delaware Verbal Morphology: A Descriptive and Comparative Study*. New York: Garland.

156

Goddard, I. (1987). Leonard Bloomfield's descriptive and comparative studies of Algonquian. In R. A. Hall and K. Koerner (Eds.), *Leonard Bloomfield: Essays on his Life and Work*, pp. 179–217. John Benjamins.

Gold, E. M. (1967). Language identification in the limit. *Information and Control 10*, 447–474.

Goldrick, M. and P. Smolensky (1999). Opacity, turbid representations, and output-based explanation. In *Workshop on the Lexicon in Phonetics and Phonology*.

Goldsmith, J., Y. Hu, I. Matveeva, and C. Sprague (2005). A heuristic for morpheme discovery based on string edit distance. Technical report, University of Chicago.

Goldwater, S. and M. Johnson (2003). Learning OT constraint rankings using a Maximum Entropy model. In J. Spenader, A. Eriksson, and Ö. Dahl (Eds.), *Proceedings of the Stockholm Workshop on 'Variation within Optimality Theory'*, pp. 111–120.

Golénia, B., S. Spiegler, and P. Flach (2009). Unsupervised morpheme discovery with ungrade. In *Multilingual Information Access Evaluation I: Text Retrieval Experiments*. Springer.

Gorman, K. (2012). Exceptions to rhotacism. In *CLS 48*. University of Chicago.

Gouskova, M. (2003). *Deriving Economy: Syncope in Optimality Theory*. Ph. D. thesis, University of Massachusetts, Amherst.

Gouskova, M. (2012). Unexceptional segments. *Natural Language and Linguistic Theory 30*(1), 79–133.

Gouskova, M. and M. Becker (2013). Nonce words show that Russian yer alternations are governed by the grammar. *Natural Language and Linguistic Theory 31*, 735–765.

Gress-Wright, J. (2010). *Opacity and Transparency in Phonological Change*. Ph. D. thesis, University of Pennsylvania.

Häberl, C. G. (2009). *The Neo-Mandaic Dialect of Khorramshahr*. Harrassowitz Verlag.

Halle, M. and J.-R. Vergnaud (1987). *An Essay On Stress*. MIT Press.

Hanzeli, V. E. (1969). *Missionary Linguistics in New France: A Study of Seventeenth- and Eighteenth-Century Descriptions of American Indian Languages*. Mouton.

Harris, J. (1994). *English Sound Structure*. Blackwell.

Hay, J. and A. Sudbury (2005). How rhoticity became /r/-sandhi. *Language 81*, 799–823.

Hayes, B. (1995). *Metrical Stress Theory*. Chicago: The University of Chicago Press.

Hayes, B. (1999a). Phonetically driven phonology: The role of Optimality Theory and inductive grounding. In *Functionalism and Formalism in Linguistics*, Volume 1, pp. 243–285. John Benjamins.

Hayes, B. (1999b). Phonological restructuring in Yidié and its theoretical consequences. In B. Hermans and M. van Oostendorp (Eds.), *The Derivational Residue in Phonological Optimality Theory*, pp. 175–205. John Benjamins.

Hayes, B. (2004). Phonological acquisition in optimality theory: The early stages. In R. Kager, J. Pater, and W. Zonneveld (Eds.), *Fixing Priorities: Constraints in Phonological Acquisition*. Cambridge University Press.

Hayes, B., R. Kirchner, and D. Steriade (Eds.) (2004). *Phonetically-Based Phonology*. Cambridge University Press.

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

Hayes, B., K. Zuraw, P. Siptar, and Z. Londe (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language 85*, 822–863.

Heinz, J. (2007). *Inductive Learning of Phonotactic Patterns*. Ph. D. thesis, University of California, Los Angeles.

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

Hermans, B. and M. van Oostendorp (Eds.) (1999). *The Derivational Residue in Phonological Optimality Theory*. John Benjamins.

Hockett, C. (1948). Potawatomi I: Phonemics, morphophonemics and morphological survey. *International Journal of American Linguistics 14*(1), 1–10.

Hoijer, H. (1933). *Tonkawa: An Indian Language of Texas*, Volume 3 of *Handbook of American Indian Languages*. Columbia University Press.

Hoijer, H. (1946). Tonkawa: An Indian language of Texas. In C. Osgood (Ed.), *Linguistic Structures of Native America*, Volume 6 of *Publications in Anthropology*. Viking Fund.

Hyman, L. (1976). Phonologization. In A. Juilland (Ed.), *Linguistic Studies Offered to Joseph Greenberg on the Occasion of his Sixtieth Birthday*, pp. 407–418. Anma Libri.

Inkelas, S., C. O. Orgun, and C. Zoll (1997). The implications of lexical exceptions for the nature of grammar. In *Constraints and Derivations in Phonology*. Clarendon Press.

Isačenko, A. (1970). East Slavic morphophonemics and the treatment of the jers in Russian: A revision of Havlík's law. *International Journal of Slavic Linguistics and Poetics 13*, 73–124.

Ito, J. and A. Mester (1992 [2003]). Weak layering and word binarity. In T. Honma, M. Okazaki, T. Tabata, and S.-i. Tanaka (Eds.), *A New Century of Phonology and Phonological Theory: A Festschrift for Professor Shosuke Haraguchi on the Occasion of his Sixtieth Birthday*, pp. 26–65. Tokyo: Kaitakusha.

Jackson, K. (1953). *Language and History in Early Britain: A Chronological Survey of the Brittonic Languages First to Twelfth Century A.D.* Edinburgh University Press.

Jacobs, H. (1989). *Nonlinear studies in the historical phonology of French*. Ph. D. thesis, University of Nijmegen.

Jacobs, H. (2004). Rhythmic vowel deletion in OT: Syncope in Latin. *Probus 16*, 63–90.

Jardine, A., J. Chandlee, R. Eyraud, and J. Heinz (2014). Very efficient learning of structured subspaces of subsequential functions from positive data. In *Proceedings of the 12th International Conference on Grammatical Inference*.

Jarosz, G. (2006). *Rich Lexicons and Restrictive Grammars: Maximum Likelihood Learning in Optimality Theory*. Ph. D. thesis, Johns Hopkins.

Kager, R. (1997). Rhythmic vowel deletion in Optimality Theory. In I. Roca (Ed.), *Derivations and Constraints in Phonology*, pp. 463–499. Oxford University Press.

Kager, R. (1999). Surface opacity of metrical structure in Optimality Theory. In B. Hermans and M. van Oostendorp (Eds.), *The Derivational Residue in Phonological Optimality Theory*, pp. 207–245. John Benjamins.

Kager, R. (2001). Rhythmic directionality by positional licensing. In *Fifth HIL Phonology Conference*. University of Potsdam.

Kavitskaya, D. and P. Staroverov (2008). Opacity in Tundra Nenets. In N. Abner and J. Bishop (Eds.), *WCCFL 27*, pp. 274–282. Cascadilla Proceedings Project.

Kawahara, S. (2002). Similarity among variants: Output-variant correspondence. Master's thesis, International Christian University, BA thesis.

Kaye, J. (1973). Odawa stress and related phenomena. In *Odawa Language Project: Second Report*. University of Toronto.

Kaye, J. (1974a). Morpheme structure constraints live! In *Montreal Working Papers in Linguistics*, Volume 3, pp. 55–62. McGill University.

Kaye, J. (1974b). Opacity and recoverability in phonology. *Canadian Journal of Linguistics 19*(2), 134–149.

Kaye, J. and B. Nykiel (1979). Loan words and abstract phonotactic constraints. *Canadian Journal of Linguistics 24*(2), 71–93.

Keating, P. (1985). Universal phonetics and the organization of grammars. In V. Fromkin (Ed.), *Phonetic Linguistics: Essays in honor of Peter Ladefoged*. Academic Press.

Keating, P. (1996). The phonology-phonetics interface. In U. Kleinhenz (Ed.), *Interfaces in Phonology*, pp. 262–278. Akademie Verlag.

Kennedy, R. (2003). *Confluence in Phonology: Evidence from Micronesian Reduplication*. Ph. D. thesis, University of Arizona.

Kenstowicz, M. (1980). Notes on Cairene Arabic syncope. In M. Kenstowicz (Ed.), *Studies in the Linguistic Sciences*, Volume 10, pp. 39–53. University of Illinois, Urbana-Champaign.

Kenstowicz, M. (1996). Base identity and uniform exponence: Alternatives to cyclicity. In J. Durand and B. Laks (Eds.), *Current Trends in Phonology: Models and Methods*, pp. 363–393. University of Salford Publications.

Kenstowicz, M. and C. Kisseberth (1977). *Topics in Phonological Theory*. University of Illinois.

Kenstowicz, M. and C. Kisseberth (1979). *Generative Phonology*. Academic Press.

Kim, Y. J. (2015). *6-month-olds' Segmentation and Representation of Morphologically Complex Words*. Ph. D. thesis, University of California, Los Angeles.

Kimper, W. (2011). Locality and globality in phonological variation. *Natural Language and Linguistic Theory 29*, 423–465.

King, R. (1969). *Historical Linguistics and Generative Grammar*. Prentice-Hall.

King, R. (1976). *The History of Final Devoicing in Yiddish*. Indiana University Linguistics Club.

King, R. (1988). A problem of vowel lengthening in Early New High German. *Monatshefte 80*, 22–31.

Kiparsky, P. (1965). *Phonological Change*. Ph. D. thesis, MIT.

Kiparsky, P. (1968a). *How Abstract Is Phonology?* Indiana University Linguistics Club.

161

Kiparsky, P. (1968b). Linguistic universals and linguistic change. In E. Bach and R. Harms (Eds.), *Universals in Linguistic Theory*. Holt, Rinehart and Winston.

Kiparsky, P. (1971). Historical linguistics. In W. O. Dingwall (Ed.), *A Survey of Linguistic Science*, pp. 576–649. University of Maryland Press.

Kiparsky, P. (1973). Abstractness, opacity and global rules. In O. Fujimora (Ed.), *Three Dimensions of Linguistic Theory*, pp. 57–86. Tokyo: Tokyo Institute for Advanced Studies of Language.

Kiparsky, P. (1978). Analogical change as a problem for linguistic theory. In B. B. Kachru (Ed.), *Linguistics in the Seventies: Directions and Prospects*, Volume 8, pp. 77–96. University of Illinois, Urbana-Champaign.

Kiparsky, P. (1988). Phonological change. In F. Newmeyer (Ed.), *Linguistics: The Cambridge Survey*, Volume 1, Chapter 14, pp. 363–415. Cambridge University Press.

Kiparsky, P. (1995). The phonological basis of sound change. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory*. Blackwell.

Kiparsky, P. (2000). Opacity and cyclicity. *The Linguistic Review 17*, 351–367.

Kiparsky, P. (2006). Amphichronic linguistics vs. Evolutionary Phonology. *Theoretical Linguistics 32*, 217–236.

Kiparsky, P. (2008). Universals constrain change, change results in typological generalizations. In *Linguistic Universals and Language Change*. Oxford University Press.

Kiparsky, V. (1979). *Russian Historical Grammar*. Ardis.

Kirchner, R. (1995, May). Going the distance: Synchronic chain shifts in Optimality Theory. UCLA manuscript, available as ROA 66.

Kisseberth, C. (1970). On the functional unity of phonological rules. *Linguistic Inquiry 1*(3), 291–306.

Koch, J. (1995). The conversion and the transition from Primitive to Old Irish. *Emania: Bulletin of the Navan Research Group 13*, 39–50.

Kuroda, S.-Y. (1967). *Yawelmani Phonology*. MIT Press.

Lightfoot, D. (1999). *The Development of Language: Acquisition, Change and Evolution*. Blackwell Publishers, Inc.

Linzen, T., S. Kasyanenko, and M. Gouskova (2013). Lexical and phonological variation in Russian prepositions. *Phonology 30*(3).

Lunt, H. (1980). On "akanje" and linguistic theory. *Harvard Ukrainian Studies 3/4*, 595–608.

Lynch, J. (2001). Article accretion and article creation in Southern Oceanic. *Oceanic Linguistics 40*(2), 224–246.

Magri, G. Idempotency in Optimality Theory. Ms, CNRS and Utrecht University.

Magri, G. (2013). The complexity of learning in Optimality Theory and its implications for the acquisition of phonotactics. *Linguistic Inquiry 44*(3), 433–468.

Magri, G. (2015a). Idempotency and chain shifts. In K.-m. Kim (Ed.), *Proceedings of the 33rd annual West Coast Conference in Formal Linguistics, WCCFL*. Cascadilla.

Magri, G. (2015b, January). Output-drivenness and partial phonological features. CNRS manuscript.

Malone, J. (1997). Modern and classical Mandaic phonology. In A. S. Kaye (Ed.), *Phonologies of Asia and Africa*, Volume 1, Chapter 10, pp. 141–159. Eisenbrauns.

Martin, A. (2007). *The Evolving Lexicon*. Ph. D. thesis, University of California, Los Angeles.

McCarthy, J. (1991). Synchronic rule inversion. In *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*, pp. 192–207.

McCarthy, J. (1993). A case of surface constraint violation. *Canadian Journal of Linguistics 38*, 169–195.

McCarthy, J. (2003). Comparative markedness. *Theoretical Linguistics 29*, 1–51.

McCarthy, J. (2007a). *Hidden Generalizations: Phonological Opacity in Optimality Theory*. Equinox Publishing.

McCarthy, J. (2008). The serial interaction of stress and syncope. *Natural Language and Linguistic Theory 26*, 499–546.

McCarthy, J. (2010). An introduction to Harmonic Serialism. *Language and Linguistics Compass 4*(10), 1001–1018.

McCarthy, J. J. (2007b). Slouching toward optimality: Coda reduction in OT-CC. *Journal of the Phonological Society of Japan 7*.

McMahon, A. (2000). *Lexical Phonology and the History of English*. Cambridge University Press.

McManus, D. (1983). A chronology of the Latin loan-words in Early Irish. *Ériu 34*, 21–71.

Merchant, N. N. (2008). *Discovering Underlying Forms: Contrast Pairs and Ranking*. Ph. D. thesis, Rutgers.

Minkova, D. (1991). *The History of Final Vowels in English: The Sound of Muting*. Mouton de Gruyter.

Moreton, E. (2008). Analytic bias and phonological typology. *Phonology 25*.

Moreton, E. and P. Smolensky (2002). Typological consequences of local constraint conjunction. In L. Mikkelsen and C. Potts (Eds.), *WCCFL 21*. Cascadilla Press.

Munshi, S. and M. Crowhurst (2012). Weight sensitivity and syllable codas in Srinagar Koshur. *Journal of Linguistics 48*(2), 427–472.

Nichols, J. and E. Nyholm (1995). *A Concise Dictionary of Minnesota Ojibwe*. University of Minnesota Press.

Nishimura, K. (2010). Patterns of vowel reduction in Latin: Phonetics and phonology. *Historische Sprachforschung 123*, 217–257.

Nishimura, K. (2012). Vowel reduction and deletion in Sabellic: A syncronic and diachronic interface. In B. Whitehead, T. Olander, B. Olsen, and J. Rasmussen (Eds.), *The Sound of Indo-European: Phonetics, Phonemics, and Morphophonemics*, Volume 4 of *Copenhagen Studies in Indo-European*, pp. 381–398. Museum Tusculanum Press.

Niyogi, P. (2006). *The Computational Nature of Language Learning and Evolution*. MIT Press.

Ohala, J. (1989). Sound change is drawn from a pool of synchronic variation. In L. E. Breivik and E. H. Jahr (Eds.), *Language Change: Contributions to the Study of its Causes*, pp. 173–198. Mouton de Gruyter.

Ohala, J. (1992). What's cognitive, what's not, in sound change. In G. Kellermann and M. Morrissey (Eds.), *Diachrony within Synchrony: Language History and Cognition Papers from the International Symposium at the University of Duisburg*, pp. 309–355. Frankfurt am Main.

Ohala, J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical Linguistics: problems and perspectives*, pp. 237–278. Longman.

Ohala, M. (1977). The treatment of phonological variation: An example from Hindi. *Lingua 42*, 161–176.

Overall, S. (2007). *A Grammar of Aguaruna*. Ph. D. thesis, La Trobe University.

Padgett, J. (2001). Contrast dispersion and Russian palatalization. In *The role of Speech Perception in Phonology*. Academic Press.

Padgett, J. and M. Tabain (2005). Adaptive Dispersion Theory and phonological vowel reduction in Russian. *Phonetica 62*, 14–54.

Pater, J. (2010). Morpheme-specific phonology: Constraint indexation and inconsistency resolution. In S. Parker (Ed.), *Phonological Argumentation: Essays on Evidence and Motivation*, Chapter 5. Equinox.

Payne, D. (1990). Accent in Aguaruna. In D. Payne (Ed.), *Amazonian Linguistics: Studies in Lowland South America*, pp. 161–184. University of Texas Press.

Pesetsky, D. (1979, March). Russian morphology and lexical theory. M.S. MIT.

Piggott, G. L. (1974 [1980]). *Aspects of Odawa Morphophonemics*. Garland.

Piggott, G. L. (1983). Extrametricality and Ojibwa stress. In *McGill Working Papers in Linguistics*, Volume 1, pp. 80–118. McGill University.

Piggott, G. L. and J. Kaye (1973). Odawa language project: Second report. Technical report, University of Toronto.

Piggott, G. L., J. Kaye, and K. Tokaichi (1971). Odawa language project: First report. Technical report, University of Toronto.

Pope, M. (1952). *From Latin to Modern French with Especial Consideration of Anglo-Norman Phonology and Morphology* (2 ed.). University of Manchester Press.

Prince, A. (2002, February). Entailed ranking arguments. Technical report, Rutgers University Department of Linguistics and Center for Cognitive Science.

Prince, A. and A. Brasoveanu (2010). The formal structure of ranking arguments in OT. Manuscript, Rutgers University and UC Santa Cruz.

Prince, A. and P. Smolensky (1993 [2004]). Optimality Theory: Constraint interaction in generative grammar. Technical report, University of Colorado at Boulder.

Prince, A. and B. Tesar (2004). Learning phonotactic distributions. In R. Kager, J. Pater, and W. Zonneveld (Eds.), *Constraints in Phonological Acquisition*, pp. 245–291. Cambridge University Press.

Rhodes, R. (1976). *The Morphosyntax of the Central Ojibwa Verb*. Ph. D. thesis, University of Michigan.

Rhodes, R. (1985a). *Eastern Ojibwa-Chippewa-Ottawa Dictionary*. Mouton.

Rhodes, R. (1985b). Lexicography and Ojibwa Vowel Deletion. *The Canadian Journal of Linguistics 30*(4), 453–471.

Richards, N. (1997). Leerdil yuujmen bana yanangarr (old and new Lardil). In *MIT Occasional Papers in Linguistics: Papers on Australian Languages*, Volume 13. MIT Working Papers in Linguistics.

Rickard, P. (1989). *A History of the French Language* (2 ed.). Unwin Hyman.

Riggle, J. (2004). *Generation, Recognition, and Learning in Finite State Optimality Theory*. Ph. D. thesis, University of California, Los Angeles.

Riggle, J. (2006a). Infixing Reduplication in Pima and its theoretical consequences. *Natural Language and Linguistic Theory*.

Riggle, J. (2006b). Using entropy to learn OT grammars from surface forms alone. In *Proceedings of the 25th West Coast Conference on Formal Linguistics*.

Ristad, E. and P. N. Yianilos (1996). Learning string edit distance. Princeton University, Department of Computer Science, Research Report CS-TR-532-96.

Ritter, E. and S. T. Rosen (2005). Agreement without A-positions: Another look at Algonquian. *Linguistic Inquiry 36*(4), 648–660.

Ritter, E. and S. T. Rosen (2009). Animacy in Blackfoot: Implications for event structure and clause structure. In *Syntax, Lexical Semantics, and Event Structure*. Oxford University Press.

Samek-Lodovici, V. and A. Prince. Optima. ROA-363.

Sanders, R. N. (2003). *Opacity and Sound Change in the Polish Lexicon*. Ph. D. thesis, University of California, Santa Cruz.

Selkirk, E. O. (1995). The prosodic structure of function words. In *Papers in Optimality Theory II (University of Massachusetts Occasional Papers in Linguistics)*, pp. 439–470. GLSA Publications.

Sóskuthy, M. (2013). Analogy in the emergence of intrusive-r in English. *English Language and Linguistics 17*, 55–84.

Stampe, D. (1973). *How I Spent my Summer Vacation: A Dissertation on Natural Phonology*. Ph. D. thesis, University of Chicago.

Staroverov, P. (2006). Vowel deletion and stress in Tundra Nenets. In B. Gyuris (Ed.), *Proceedings of the First Central European Student Conference in Linguistics*. Hungarian Academy of Sciences.

Stump, G. T. (2006). Heteroclisis and paradigm linkage. *Language 82*, 279–322.

Summer Institute of Linguistics (2005). *Diccionario Tepehuano de Santa María Ocotán, Durango*. Summer Institute of Linguistics, Mexico.

Tesar, B. (2013). *Output-Driven Phonology: Theory and Learning*. Cambridge University Press.

Tesar, B. and A. Prince (2003 [2007]). Using phonotactics to learn phonological alternations. In J. Cihlar, A. Franklin, D. Kaiser, and I. Kimbara (Eds.), *Papers from the 39th Annual Meeting of the Chicago Linguistic Society*, Number 2, pp. 209–237.

Tesar, B. and P. Smolensky (1993). Learnability of Optimality Theory. ROA-2.

Tesar, B. and P. Smolensky (1998). Learnability in Optimality Theory. *Linguistic Inquiry 29*(2), 229–268.

Tesar, B. and P. Smolensky (2000). *Learnability in Optimality Theory*. MIT Press.

Thurneysen, R. (1946). *Grammar of Old Irish*. The Dublin Institute for Advanced Studies.

University of Minnesota's Dept. of American Indian Studies and University Libraries (2012). The Ojibwe People's Dictionary.

Uwarai Yagkug, A. and I. Paz Suikai (1998). *Diccionario Aguaruna Castellano: Awajún Chícham Apáchnaujai*. Centro Amazónico de Antropología y Aplicación Práctica.

Valentine, R. (1994). *Ojibwe Dialect Relations*. Ph. D. thesis, University of Texas, Austin.

Valentine, R. (2001). *Nishnaabemwin Reference Grammar*. Toronto: University of Toronto Press, Inc.

Vennemann, T. (1972). Phonetic analogy and conceptual analogy. In *Schuchhardt, the Neogrammarians, and the transformational theory of phonological change: four essays by Hugo Schuchhardt*, pp. 183–204. Athenaum.

White, J. (2013). *Bias in Phonological Learning: Evidence from Saltation*. Ph. D. thesis, University of California, Los Angeles.

White, J. (2014). Evidence for a learning bias against saltatory phonological alternations. *Cognition 130*, 96–115.

Willett, E. (1982). Reduplication and accent in Southeastern Tepehuan. *International Journal of American Linguistics 48*(2), 168–184.

Willett, T. L. (1991). *A Reference Grammar of Southeastern Tepehuan*. Summer Institute of Linguistics.

Williams, A. (1991). *The Dog's Children: Anishinaabe Texts Told by Angeline Williams*. Winnipeg: The University of Manitoba Press.

Wilson, C. (2001). Consonant cluster neutralisation and targeted constraints. *Phonology 18*(1), 147–197.

Winter, W. (1971). Formal frequency and linguistic change: Some preliminary comments. *Folia Linguistica*.

Wipio Deicat, G. (1996). *Diccionario Aguaruna-Castellano Castellano-Aguaruna*, Volume 39 of *Serie Lingüística Peruana*. Lima, Peru: Instituto Lingüístico de Verano.

Yearley, J. (1995). Jer vowels in Russian. In J. Beckman, L. W. Dickey, and S. Urbanczyk (Eds.), *Papers in Optimality Theory II (University of Massachusetts Occasional Papers in Linguistics)*, pp. 533–571. GLSA Publications.

Yu, A. (2004). Explaining final obstruent voicing in Lezgian: Phonetics and voicing. *Language 80*, 73–97.

Zaliznjak, A. A. (1977). *Grammatečeski slovar' russkogo jazyka [A Grammatical Dictionary of the Russian language]*. Russkij Jazyk.

Zuraw, K. (2007, December). Projecting new forms from old in Palauan. UCLA ms.