Phonological Data and Analysis Volume X, Article X: pages (year)
https://doi.org/00.0000/nn.0.00

Received –; revised –; accepted –. ©date author
Published by the Linguistic Society of America with permission of the authors under a CC BY 3.0 license.

Evidence for prosodic correspondence in the vowel alternations of Tgdaya Seediq

author

This paper brings new evidence for PROSODIC CORRESPONDENCE, where prosodic units (e.g. main-stressed nuclei and prominent syllables) of morphologically related forms are compared. Since prosodic correspondence was formalized in Crosswhite's (1998) analysis of Chamorro, it has received almost no empirical discussion. I argue that Tgdaya Seediq (Austronesian, Atayalic) has vowel alternations which should be analyzed using prosodic correspondence. In Seediq, stem and suffixed forms tend to share the same stressed syllable nucleus. This VOWEL MATCHING pattern cannot be explained as surface harmony, but it can be explained as the result of a constraint enforcing vowel identity of main-stressed nuclei in morphologically related forms. Unlike the categorical alternations analyzed by Crosswhite (1998), Seediq vowel matching is gradient and only emerges on a statistical level. Nevertheless, prosodic correspondence appears to be active in the synchronic grammar of Seediq; in a production experiment, speakers applied vowel matching to novel forms, and even over-generalized it to environments not predicted by lexical statistics. Vowel matching is modeled in Maximum Entropy Harmonic Grammar (Goldwater & Johnson 2003), a stochastic variant of OT. I use prosodic correspondence to model vowel matching, and Zuraw's (2000) dual listing approach to capture the discrepancy between lexical and experimental results.

Keywords: phonology;vowel alternations;vowels;statistical learning;Seediq; prosodic correspondence

1 Introduction

Morphologically related forms tend to be similar with respect to one another. In Optimality Theory (OT; Smolensky 1986; Prince & Smolensky 1993, 2004), this similarity effect has been formalized in various ways, including Correspondence (McCarthy & Prince 1995; Benua 1995), Paradigm Uniformity (Steriade 2000), and Uniform Exponence (Kenstowicz 1995).[1] These theories generally assume the comparison of segmental elements that are linearly aligned, as schematized in (1).

(1) *Segmental comparison of a stem and its suffixed form*

$$\begin{array}{ccccc} S_1 & S_2 & S_3 & & \textsc{stem} \\ | & | & | & & \\ \mathbb{S}_1 & \mathbb{S}_2 & \mathbb{S}_3 & \text{-}\textsc{suff} & \textsc{stem+suffix} \end{array}$$

In her work on Chamorro, however, Crosswhite (1998) finds evidence for a different type of comparison, where *prosodic*, rather than segmental, elements of morphologically related output forms are compared. As will be elaborated on in Section 2, this concept is formalized as PROSODIC CORRESPONDENCE, and involves

[1] Note that in Stratal OT, paradigm similarity effects fall out naturally from cyclicity, and can be treated using just IO-correspondence constraints (Kiparsky 2000, 2015). In this paper, I assume a parallel OT framework, but the main issue addressed in this paper, which deals with prosodic vs. segmental correspondence, is relevant to both parallel and stratal OT.

comparison of units such as the most prominent syllable nodes of each stem, or the main-stressed nucleus of each stem. To the author's knowledge, no other examples of prosodic correspondence have been documented. Building on Crosswhite's findings, the purpose of this paper is to come back to the idea of prosodic correspondence, with novel evidence from the vowel alternations of Tgdaya Seediq.

Tgdaya Seediq is a dialect of Seediq (Austronesian, Atayalic). In Tgdaya Seediq (henceforth Seediq), stress is always penultimate, and suffixation shifts stress one syllable to the right. Seediq has two fundamental processes of stress-related vowel reduction, which result in alternations between forms of a paradigm. These processes are introduced here, and detailed in Section 3.

The first process, called pretonic reduction, causes all vowels to be reduced to [u] pretonically. As seen in (2), pretonic reduction causes the first vowel of the stem to surface as [u] in the suffixed form.

(2) *Examples of pretonic vowel redution in Seediq*

STEM	SUFFIXED	UR	
ˈbeliŋ	buˈliŋ-an	/beliŋ/	'cave'
ˈbiciq	buˈciq-an	/biciq/	'few'
ˈcapaŋ	cuˈpaŋ-an	/capaŋ/	'thick cloth'

The second process, which is directly relevant to the current paper, is post-tonic vowel reduction. This process, illustrated below in (3), causes mid vowels (/e, o/) to reduce to [u] post-tonically. As a result, the final /e/ and /o/ of the stem will surface as [u] in non-suffixed allomorphs (e.g. the isolation stem). This results in [u]∼[e] and [u]∼[o] alternations between stem and suffixed forms of a paradigm.

(3) *Examples of vowel alternation in Seediq verb paradigms*

	STEM	SUFFIXED	UR	
(a)	ˈp**e**mux	puˈm**e**xan	/pemex/	'hold'
(b)	ˈk**o**duŋ	kuˈd**o**ŋ-an	/kodoŋ/	'hook'

Crucially, the current study finds a strong propensity for alternations of the post-tonic [u] ([u]∼[e] and [u]∼[o]) to result in VOWEL MATCHING, where the stressed vowels of stem and suffixed allomorphs match. For example, in (3a), the stressed vowels of the isolation stem and suffixed forms, indicated in boldface, are both [e].

Vowel matching is not just present as a statistical tendency in the lexicon. In a production experiment, speakers were asked to produce novel suffixed forms, given stimuli with neutralized post-tonic vowels. This methodology, which tests speakers' application of alternations to novel items, has been shown to reflect speakers' phonological knowledge of gradient patterns in the lexicon (e.g. Ernestus & Baayen 2003; Becker, Ketrez & Nevins 2011). In the current experiment, speakers were found to successfully 'undo' post-tonic vowel neutralization, and productively apply vowel alternations in a way that resulted in vowel matching. Moreover, speakers' application of vowel matching actually extended beyond environments predicted by lexical distributions.

Seediq VOWEL MATCHING resembles vowel harmony as well as copy epenthesis, where the quality of an epenthetic vowel depends on the quality of a vocalic neighbor (Stanton & Zukoff 2018). However, traditional analyses for both harmony and copy epenthesis, such as autosegmental feature spreading (e.g. Clements 1986; Kawahara 2007), segmental correspondence (e.g. Kitto & De Lacy 1999), or agreement by correspondence (Rose & Walker 2004), do not provide a straightforward analysis for the Seediq data. This is because Seediq does not have surface vowel agreement. For example, returning to example (3a), because of post-tonic and pretonic vowel reduction, neither allomorphs ([ˈpemux], [puˈmexan]) show surface vowel matching. Instead, vowel matching is only observed by comparing prosodic positions (specifically the nucleus of stressed syllables) across related surface forms.

As such, I will argue that Seediq vowel matching is evidence for PROSODIC CORRESPONDENCE (Crosswhite 1998). Specifically, vowel matching is the result of an output-output correspondence relationship between the stressed nuclei of morphologically related forms. In this paper, I will motivate the need for prosodic correspondence in Seediq, and provide an OT analysis for Seediq vowel matching.

The rest of this paper is organized as follows. First, Section 2 introduces prosodic correspondence, and explains how it differs from segmental correspondence. Section 3 describes relevant Seediq vowel alternations. In Section 4, I conduct a survey of a Seediq corpus, and demonstrate that there is a strong statistical tendency towards vowel matching in specific environments. In Section 5, I show that in a production experiment, speakers actually *overgeneralize* and extend vowel matching beyond environments provided by the lexicon. Following this, Section 6 provides an analysis of Seediq vowel matching set in OT. Finally, in Section 7, I briefly discuss whether there is a historical basis for Seediq vowel matching.

2 Segmental vs. prosodic correspondence

In OT, phonological similarity effects have most commonly been formalized in terms of McCarthy & Prince's (1995) theory of Correspondence. Correspondence captures the idea that elements of a given form are paired with elements of related forms. The tendency for related forms to be similar can be captured using Identity (Ident) constraints, which assign violations when corresponding elements are not the same for a feature **F**.

There are various types of correspondence relationships, including input-output, base-reduplicant, and output-output. For example, given the Seediq stem-suffix pair [ˈpemux]∼[puˈmexan] /pemex/, there are at least two possible correspondence relations, illustrated in (4). First, the surface form [ˈpemux] is in an Input-Output (IO) correspondence relation with its UR /pemex/. Morphologically related surface forms [ˈpemux] and [puˈmexan] are related via Output-Output (OO) correspondence. In addition, although not depicted here, the suffixed form [puˈmexan] also corresponds to the UR (via IO correspondence).

(4) *Segmental correspondence relations: Seediq example*

	/ p e m e x /	'hold'
IO	| | | | |	
	[ˈp e m u x]	
OO	| | | | |	
	[p u ˈm u x -an]	

The correspondences shown in (4) assume linear mappings along the segmental dimension. However, instead of comparing segmental elements, it is also possible to compare prosodic slots under a correspondence-based theory. It is well-recognized that prosodic positions are relevant in correspondence relationships has been explored in various work. Typologically, certain prosodic positions, such as stressed syllables or foot heads, are more salient. Based on this, Kenstowicz (1995) has suggested that identity with respect to salient prosodic positions can rank above identity for typically correspondent segments.

prosodic correspondence takes the status of prosodic units even further, and compares prosodic units of related forms, even if they are not linearly and segmentally related. These types of constraints were first brought up in McCarthy (1995), and fleshed out in more detail by Crosswhite (1998). The rest of this section and subsequent analysis is adapted from Crosswhite's formalization of prosodic correspondence.[2]

One possible prosodic correspondence relation is the relation between stressed syllable nuclei of morphologically related words. In Seediq, for example, this would involve the correspondence relation shown as the solid line in (5). If this type of correspondence exists, the same vowel in a surface stem could simultaneously correspond to two separate vowels in the derived forms–it's segmental correspondent (indicated with a dotted line), and the prosodic correspondent (the stressed syllable nucleus, indicated with a solid line).

[2] The current analysis could be adapted for other theories of phonological similarity, as Paradigm Uniformity (Steriade 2000) and Uniform Exponence (Kenstowicz 1995). For simplicity, I follow Crosswhite's approach and adopt correspondence constraints

(5) *Prosodic correspondence relations: Seediq example*

OO
[(ˈp e) m u x] ‘hold’
[p u (ˈm e) x an]

Building off of Chung’s (1983) work on Chamorro, Crosswhite (1998) finds evidence for this type of simultaneous correspondence in Charmorro. Specifically, she uses a correspondence relation between the nuclei of main-stressed syllables to explain a length alternation in Chamorro. In Chamorro, stressed syllables must be heavy; for open syllables, this is usually achieved by lengthening the vowel (e.g. /finu-ɲa/→[fi.ˈnuː.ɲa] ‘smoother’). In a subset of suffixed forms, however, a heavy stressed syllable is achieved by geminating the suffix, instead of lengthening the stem vowel (e.g. /lebblu-ɲa/→[leb.ˈbloɲ.ɲa]).

Crosswhite (1998) argues that gemination is used to ensure that stressed vowel nuclei of related forms share the same moraic length. Gemination ensures that if the base allomorph has main stress on a closed syllable with a short vowel, the main-stressed syllable of derived forms will also have a short vowel. Examples of gemination are given in (6). In (6a-c), the suffix /-ɲa/ surfaces as [ɲɲa], allowing the primary stress to fall on a closed syllable in the suffixed form. Similarly, in (6d-f), the suffix /-mu/ geminates to [-mmu], so that primary stress will fall on a closed syllable. Main-stressed syllables are indicated in boldface.

(6) *Gemination of suffixes in Chamorro* (Chung 1983).[3]

	STEM		STEM+ /ɲa/	
a.	ˈ**leb**.blu	‘short’	leb.ˈ**blo**ɲ.ɲa	‘shorter’
b.	ˈ**daŋ**.ku.lu	‘big’	ˌdaŋ.ku.ˈ**lo**ɲ.ɲa	‘bigger’
c.	ˈ**maj**.pi	‘hot’	maj.ˈ**pe**ɲ.ɲa	‘hotter’
	STEM		STEM+ /mu/	
d.	ˈ**tom**.mu	‘knee’	tom.ˈ**mom**.mu	‘his knee’
e.	ˈ**kan**.ta	‘song’	kan.ˈ**tam**.mu	‘your song’
f.	ˈ**swed**.du	‘salary’	swed.ˈ**dom**.mu	‘your salary’

This prosodic correspondence effect is formalized using the the constraint NUCLEUS-IDENT-OO(μ), defined in (7). This constraint enforces the correspondence relationship illustrated in (8), between the nucleus positions in the stressed syllables of morphologically related forms. Specifically, NUCLEUS-IDENT-OO(μ) requires that these positions have the same moraic length.

(7) NUCLEUS-IDENT-OO(μ): For α, a stressed nucleus of the base allomorph, and β, a stressed nucleus of a corresponding affixed form, where $\alpha\mathfrak{R}\beta$, if α is monomoraic, β must be monomoraic.

(8) *Correspondence between stressed syllable nuclei in Chamorro*

OO
[(ˈleb).blu] ‘short’
[leb.(ˈbloɲ).ɲa] ‘shorter’

NUC-IDENT-OO(μ) identifies two main-stressed nuclei and determines whether they are both associated with only one mora on the moraic tier. In my analysis, I will argue that Seediq vowel matching can be explained using the same correspondence relation, extended to vowel alternations.

[3] Some examples also involve a stress-dependent vowel height alternation. For more details on this alternation, refer to Chung (1983) and Crosswhite (1998).

Specifically, the Seediq tendency towards vowel matching will be formalized as the effect of vowel feature identity constraints between the stressed nuclei of morphologically related forms. These constraints will identify two stressed nuclei, and determine if they are associated with the same vowel features on the *segmental tier*. The constraint formulation I will use is given in (9), where F represents the relevant vowel features (e.g. [high], [back]).

(9) NUC-IDENT-OO[F]: For α, a stressed vowel of the stem and $\beta \in$, a stressed vowel of the suffixed form, where $\alpha\mathfrak{R}\beta$, if α is [γF], then β is [γF].
i.e. stressed nuclei of corresponding stem/suffixed forms should agree in value for [high].

Crosswhite's Chamorro example is categorical, such that gemination predictably occurs to preserve the moraicity of the main-stressed nucleus. In contrast, Seediq vowel matching is a gradient process. However, I will argue that vowel matching is active in the synchronic grammar of Seediq speakers; it is not only present as a strong lexical propensity in the lexicon, but was also productively extended by speakers to novel suffixed forms (see Section 5). As such, vowel matching can be modeled as the result of a gradient prosodic correspondence constraint.

The following sections will describe the empirical facts of vowel alternation in Seediq. Following this, Section 6 will lay out a detailed stochastic constraint-based analysis of Seediq vowel alternations, and demonstrate the effect of NUC-OO-IDENT[F].

3 Vowel alternations in Seediq verb paradigms

Seediq is spoken in Central and Eastern Taiwan. There are around 6500 Seediq people living in Nantou, where the Tgdaya dialect, the focus of the current study, is primarily spoken (Council of Indigenous People 2020). However, the number of fluent speakers is thought to be much fewer than this, due to high rates of language attrition.

The Seediq phoneme inventory is given in (10) and (11); where the orthography that I adopt differs from standard IPA, phonetic transcription is given in brackets. Seediq verbs are almost always inflected for voice, mood, and aspect; verbal inflection can take the form of prefixes, infixes or suffixes (Holmer 1996). These affixes are summarised in Table 1. Crucially, distributional restrictions cause there to be extensive vowel and consonant alternations between the non-suffixed and suffixed forms of a verb paradigm.[4]

During elicitation of verb paradigms (described in Section 3.1), all verbs were elicited with the /su-/, /-an/, /-un/, and /-i/ affixes. However, because the patterns reported in the paper were found to be consistent across affixes, all examples will only compare the bare stem forms (which are representative of all non-suffixed slots of the paradigm) to forms suffixed with /-an/ 'LOCATIVE FOCUS.PRES' (which are representative of all suffixed slots).

	AGENT FOCUS	LOCATIVE FOCUS	PATIENT FOCUS	INSTRU. FOCUS
PRES	-m-/mu-	-an	-un	su-
PRET	-mun-	-n-, -an	-un-	
FUT	mu(pu)-	RED-an	RED-un	
IMP		-ani	-i	

Table 1: Inflectional morphology of Seediq

[4] The current study focuses on vowel alternations, but descriptions of Seediq consonant alternations can be found in: Yang (1976); Holmer (1996); Li (1991).

(10) *Seediq consonant inventory*

Stops	*p b*	*t d*		*k g*	*q*	ʔ
Fricatives		*s*		*x*		*h*
Affricates		*c* [t͡s]				
Nasals	*m*	*n*		ŋ		
Approximants		*r* [ɾ]	*y* [j]	*w*		
Laterals		l				

(11) *Seediq vowel inventory*

i		u
e		o
	a	

3.1 Data collection

The alternations to be described in the rest of this section are based both on existing descriptive work by Yang (1976), as well as a corpus of 341 verbal paradigms.[5] These 341 paradigms were drawn from (1) the Taiwan Aboriginal e-Dictionary (Council of Indigenous Peoples 2020), and (2) fieldwork with three Seediq speakers (ages 69-78), carried out by the author in Puli Township, Nantou, Taiwan. Data was collected over the course of three weeks in July 2019. There is a high rate of language attrition in Seediq communities, such that fluent speakers are mostly above age 40, and only speakers around age 60 and above consistently use Seediq in daily conversation. As such, the speakers consulted in this study likely represent a more conservative variant of Seediq. All three consultants reported speaking Mandarin and Seediq regularly at roughly equal rates.

185 paradigms were collected from the online dictionary, and the remaining 156 paradigms were collected from native speaker consultants. Verb paradigms taken from the dictionary were confirmed with consultants, and omitted if my consultant(s) did not recognise the word, or provided conflicting inflected forms. Three forms were omitted under these criteria.

3.2 Stress-driven vowel alternations

Seediq stress is always penultimate; suffixation shifts stress rightwards (Yang 1976), giving rise to alternations such as [ˈ**bu**nuh∼buˈ**nu**han] 'wear hat'. Crucially, stress interacts with vowel neutralization processes, resulting in vowel alternations between the stem and suffixed forms of the paradigm.

Pretonically, all vowel contrasts are neutralised through reduction, deletion, or assimilation. In most cases, vowels are reduced to [u] pretonically. This is demonstrated in in (12), where the stem's initial vowel reduces to [u] when stress shifts to the second syllable in the /an/-suffixed form. Reduction to [u] occurs in 276 stems. Otherwise, if a pretonic vowel is onsetless, it is deleted, as illustrated in (13). This pattern was found for all 35 vowel-initial words in the data. Finally, the pretonic vowel will assimilate to an adjacent stressed vowel if the two are separated by [ʔ] or [h] (see (14)); 25 verbs were found to match this description. All three pretonic vowel neutralization processes are exceptionless.

(12) *Vowel reduction to [u]* (276/276)

	STEM	SUFFIX	UR	GLOSS
a.	ˈg**e**daŋ	g**u**ˈdaŋ-an	/gedaŋ/	'die'
b.	ˈb**i**ciq	b**u**ˈciq-an	/biciq/	'decrease'
c.	ˈb**a**rah	b**u**ˈrah-an	/barah/	'rare'

(13) *Onsetless vowels delete* (35/35)

a.	ˈ**a**wak	ˈwak-an	/awak/	'lead (by a leash)
b.	ˈ**e**yah	ˈyah-an	/eyah/	'come'
c.	ˈ**u**yas	ˈyas-an	/uyas/	'sing'

(14) *Vowel assimilation across [ʔ] or [h]* (25/25)

[5] A list of all 341 paradigms can be found in the supplementary materials titled 'Seediq paradigms glossed'.

a.	ˈl**e**ʔiŋ	l**i**ˈʔiŋ-an	/leʔiŋ/	'hide (an object)'
b.	ˈs**a**ʔis	s**i**ˈʔis-an	/saʔis/	'sew'

Pretonically, vowels are also optionally deleted between nasals and stops, as in (15). Although Yang (1976) describes this process as obligatory, my consultants also accepted forms where the pretonic vowel is reduced to [u], rather than deleted.

(15) *Optional vowel deletion between nasals and stops* (2/2)

	STEM	SUFFIXED	
a.	quˈn**e**dis	qunˈdis-an (∼qun**u**dis-an)	'lengthen'
b.	guˈn**a**tuk	gunˈtuk-an (∼gunuˈt**u**k-an)	'peck'

Post-tonically, similar but more restricted processes of vowel reduction are observed. Specifically, /e, o/ reduce to [u] in post-tonic closed syllables. So, for stems ending in a closed syllable with underlying /e, o, u/, the final vowel surfaces as [u] in the stem form (where it is post-tonic). However, the same vowel is stressed, and therefore surfaces faithfully, in the suffixed form. Examples of post-tonic vowel alternations are given in (16). In (16a-c), the post-tonic vowel is underlyingly /u/, and therefore non-alternating. In (16d-i), the stem has a post-tonic /e/ or /o/, resulting in [u]∼[e] and [u]∼[o] alternations.

(16) *Post-tonic reduction of /e,o/ to [u]*

a.	ˈrem**u**x ∼ ruˈm**u**xan	/remux/	'enter'	(u∼u, n=60)
b.	ˈbun**u**h∼ buˈn**u**h-an	/bunuh/	'hat'	
c.	ˈgat**u**k∼ guˈt**u**k-an	/gatuk/	'peck'	
d.	ˈpem**u**x ∼ puˈm**e**xan	/pemex/	'hold'	(u∼e , n=36)
e.	ˈhaŋ**uc** ∼ huˈŋ**e**d-an	/haŋed/	'cook, boil'	
f.	kuˈtil**u**x ∼ kutuˈl**e**x-an	/kutilex/	'hot'	
g.	doʔ**u**s ∼ doˈʔ**o**s-an	/doʔos/	'refine' (metal)'	(u∼o, n=4)
h.	toˈloʔ**u**ŋ ∼ tuloˈʔ**o**ŋ-an	/tuloʔoŋ/	'sit'	
i.	ˈkod**u**ŋ ∼ kuˈd**o**ŋ-an	/kodoŋ/	'hook'	

In addition, with the exception of /uy/, diphthongs are prohibited in post-tonic (i.e. word-final) position. /ay/ and /aw/ are respectively monophthongized to [e] and [o] as in (17a-b), while /ey/ is monophthongized to [u] as in (17c).[6]

(17) *Word-final monophthongization*

a.	ˈraŋ**e**∼ruˈŋ**ay**-an	/ranay/	'play'	(e∼ay, n=7)
b.	ˈsin**o**∼suˈn**aw**-an	/sinaw/	'to drink (alcohol)	(o∼aw,n=1)
c.	ˈdeŋ**u**∼duˈŋ**ey**-an	/deŋey/	'to dry (food)'	(u∼ey, n=12)
d.	ˈsek**u**∼suˈk**uw**-an	/seku/	'to store'	(u∼u, n=13)

Finally, there are a subset of more complex vowel alternations which interact with consonants. Specifically, as summarised in (18), final /ag/ becomes [o], and final /eg, ug/ both become [u]. These alternations are historically a result of /g/ weakening to [w] word-finally, followed by monophthongization of the resulting diphthong (Li 1981).

(18) *Alternation of final /g/*

	ALTERNATION		STEM	SUFFIXED	
a.	/ag/→[o]	(n=9)	ˈhil**o**	huˈl**ag**-a	'cover with blanket'
b.	/eg, ug/→[u]	(n=9)	ˈlih**u**	luˈh**ug**-an	'string together'
c.	/ig/→[uy]	(n=3)	ˈbar**uy**	buˈr**ig**-an	'buy/sell'

[6]Note that [o] has a limited distribution in Seediq; it surfaces post-tonically as the result of post-tonic neutralization, but there are very few stems that surface with phonemic stressed [o] as in (16c). In the current data, only four were found.

3.3 Note on forms with disyllabic suffixes

Note that there is one disyllabic suffix /-ani/ 'LOCATIVE FOCUS.IMPERATIVE'. Forms that take this suffix are assumed to not be relevant in the current discussion of vowel matching, since all vowels of the stem are neutralized when it is suffixed with /-ani/. This is demonstrated in (19). When these stems are suffixed with /-ani/, stress shifts to the first vowel of the suffix, and all of the stems' vowels are reduced to [u].

(19)		STEM	STEM-an	STEM-ani	UR	
	a.	ˈgedaŋ	g**u**ˈdaŋan	g**u**d**u**ŋani	/gedaŋ/	'die'
	b.	ˈpem**u**x	puˈm**e**xan	pumuˈxani	/pemex/	'hold'

4 Vowel matching in Seediq

In this section, I show through a survey of the Seediq lexicon that vowel matching is present as a statistical tendency in the Seediq lexicon, specifically in environments where post-tonic vowel reduction could have occurred.

4.1 Vowel matching and post-tonic vowel reduction

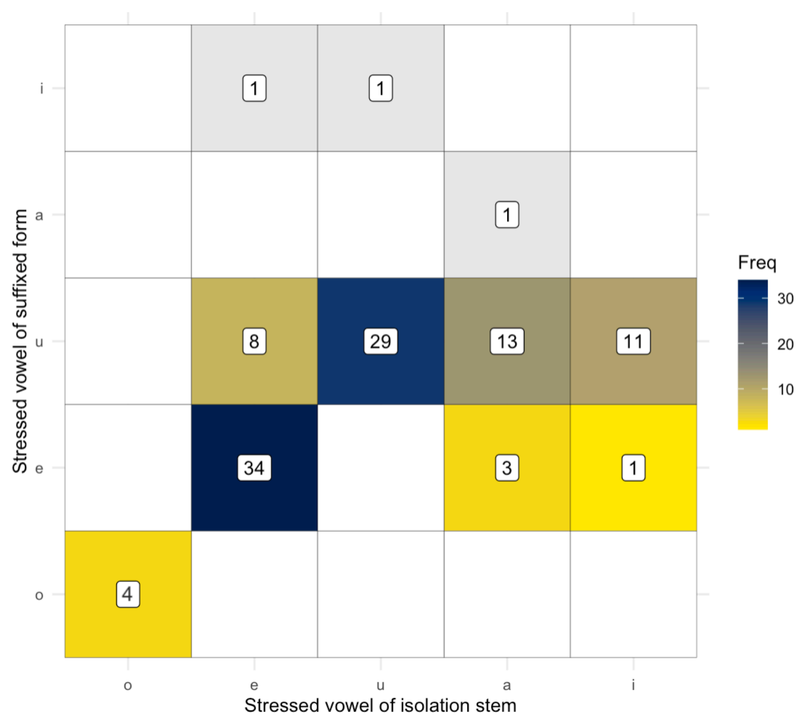

Figure 1: Stressed vowel in stem vs. suffixed form of /CVC{e,u,o}C/ words

As described above in Section 3.2, forms that are underlyingly /CVC{e,u,o}C/ undergo post-tonic reduction, and surface as [CVCuC] in the non-suffixed stem. In these surface CVCuC stems, there is a strong statistical tendency for VOWEL MATCHING, where the stressed vowel of the suffixed form matches the stressed vowel of the isolation stem. This pattern is illustrated in Fig. 1, which shows the distribution of stressed vowels in CVCuC stems. For example, the bottom-left cell (where n=4) corresponds to stem-suffix pairs like [ˈpotuk]∼[puˈtok-an], where the stressed vowel of the isolation stem, and stressed vowel of the suffixed form, are both [o].

Note that the data contains a few irregular alternations, where final /a/ or /i/ reduce to [u], resulting in

[u]∼[a,i] alternations. For example, the stem-suffix pair [ˈhuruc]∼[huˈrid-an] 'to stop suddenly' contains an irregular post-tonic [u]∼[i] alternation.[7] These are shown in the top two rows of Fig. 1 (where the stressed vowel of the suffixed form is [i] or [a]).

These exceptions aside, if the stem stressed vowel is [o], the reduced [u] always surfaces as [o] in the suffixed form (3/3, 100%). Similarly, if the stem stressed vowel is /u/, the reduced vowel will surface as [u] in the suffixed forms (29/30, 96%). For [e], there is similarly a strong tendency for vowel matching; for around 79% (34/43) of forms where the stem's stressed vowel is [e], the suffixed form's stressed vowel will also be [e]. Otherwise, if the stem stressed vowel is [a] or [i], the reduced vowel is usually non-alternating, and surfaces as [u].

This pattern was confirmed using a multinomial logistic regression model, with $V_{\text{SUFF STRESSED}}$ (stressed vowel of the suffixed allomorph) as the dependent variable. The input data was all disyllabic forms in the corpus; the predictors were V1 ([a e i o u]) and V2 ([a i u]) of the non-suffixed allomorph, as well as a binary variable VOWEL MATCHING (i.e. whether stem and suffixed allomorphs have the same stressed nucleus). If vowel matching is present in the Seediq lexicon, this last predictor should emerge as significant.

Likelihood Ratio Tests were used to confirm the significance of each predictor, and all three predictors were significant. Crucially, consistent with predictions, VOWEL MATCHING was found to be a significant predictor of the suffixed form's stressed vowel, $\chi^2(4) = 46.425$, $p < 0.0005$ ($= 2.01 \times 10^{-9}$).[8]

4.2 Vowel matching in other environments

Notably, vowel matching is only a robust pattern in stems which are susceptible to post-tonic vowel neutralizations. In other words, there is a strong propensity towards vowel matching in stems that are underlyingly /CVC{e,o,u}C/, where the final vowel surfaces as [u] in the non-suffixed allomorphs. For forms that are underlyingly /CVCaC/ or /CVCiC/, such as the ones in (20), the final vowel does not reduce in post-tonic position, and is therefore non-alternating. In this subset of forms, vowel matching is a much weaker tendency.

(20) *Examples of Seediq stems with a non-alternating final vowel*

a.	ˈbetaq∼buˈtaq-an	/betaq/	'stab, prick'
b.	ˈlaliŋ∼luˈliŋ-an	/laliŋ/	'to fish, fishhook'

This is demonstrated in Fig. 2, which plots the stressed vowel of stem forms against the stressed vowel of suffixed forms in *all* CVCVC stems, including ones where the final vowel is /a/ or /i/. Underlyingly monosyllabic stems such as /ol/ 'to follow' are excluded, since they don't show vowel alternations, and are very rare (with only 4 found in the current corpus).

As seen in Fig. 2, there is some tendency towards vowel matching, but this trend appears to be relatively weak compared to the patterns observed in Section 4.1, which looked only at contexts in which post-tonic /e,o/ are reduced (i.e. CVCuC stems). In particular, when the stressed vowel of the non-suffixed stem is /i/, the stressed vowel of the suffixed form is actually more likely to be /e/.

The stronger effect of vowel matching in CVCuC stems (relative to CVCaC and CVCiC stems) was confirmed in a logistic regression model. In this model, the dependent variable was a binary variable VOWEL MATCH. The predictors were V1 ([a e i o u]) and V2 ([a i u]) of the non-suffixed allomorph. For both predictors, the vowel [a] is the reference category. If vowel matching is a stronger tendency in CVCuC stems, then there should be an effect of V2, such that vowel matching is more likely when V2 is [u].

[7] Note that this stem-suffix pair also has a [c]∼[d] alternation, which is completely regular, and the result of a stem-final consonant neutralization process (/d/→[c]).

[8] The vowel matching tendency was also confirmed using Fisher's exact tests; the suffixed form's stressed vowel was significantly more likely to be [e] if the stem's stressed vowel was [e], $p < 0.001$ ($= 4.0 \times 10^{-15}$, odds ratio = 55). Similarly, the suffixed form's stressed vowel was significantly more likely to be [u] if the stem's stressed vowel was [u], $p < 0.001$ ($= 1.12 \times 10^{-8}$).

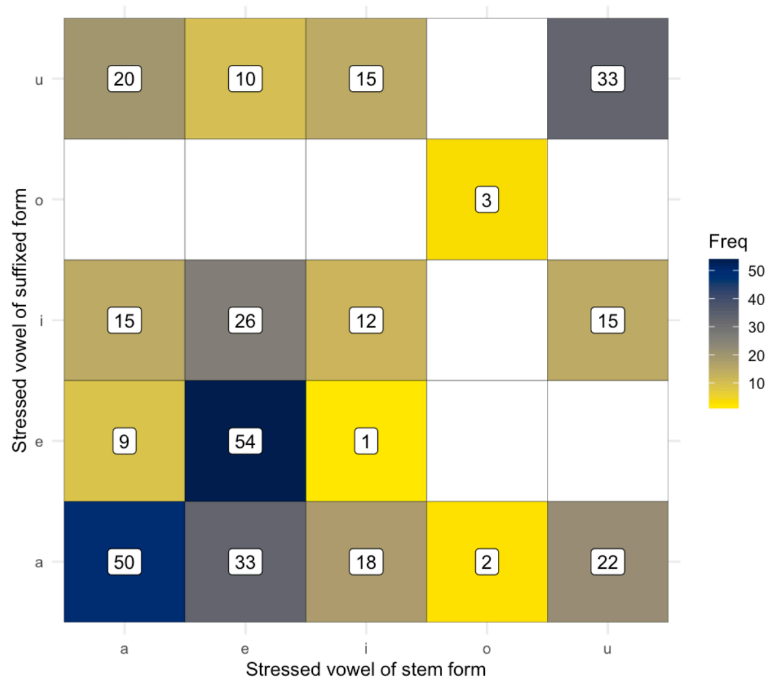

Figure 2: Distribution of stressed vowels in CVCVC stems

Model results are summarized in Table 2. Only the results for the factor V2 are reported here, but full model results are reported in Appendix A. As seen here, results are consistent with predictions; when all other predictors are held constant, the odds of vowel matching occurring was significantly higher if V2 was [u] (95% CI [1.39, 3.96], p = 0.001). Moreover, there is a weak effect for [i], such that the odds of vowel matching is actually *lower* if V2 is [i] (95% CI [0.22 - 0.91], p = 0.029)

	VOWEL MATCH		
Predictors	*Odds rations*	*CI*	*p*
(Intercept)	0.79	0.47-1.33	0.376
V2 [i]	0.45	0.22 – 0.91	0.029*
V2 [u]	**2.34**	**1.39 – 3.96**	**0.001****

Table 2: Logistic regression: Effect of V2 on vowel matching

5 Productivity of vowel matching alternations

A large body of work suggests that speaker intuitions about well-formedness are gradient (e.g. Frisch, Pierrehumbert & Broe 2004), and that speakers will apply alternations in a way which matches the frequency of alternations in the lexicon (e.g. Zuraw 2000; Ernestus & Baayen 2003; Hayes & Londe 2006; Zuraw 2010). This section discusses the results of a production experiment testing whether speakers will productively extend the vowel matching pattern that is reflected lexical statistics. As a preview, I find that speakers productively apply vowel matching alternations only to CVCuC forms (i.e. surface stems where the post-tonic vowel is [u]). In addition, speakers actually learned vowel matching non-veridically, and over-extended it to environments beyond those predicted by the lexicon.

5.1 Methodology

The experimental methodology adopted was a modified version of a nonce-word task (i.e. wug test; Berko 1958). Speakers participated in a production task, where they were given stems and asked to produce the inflected suffixed form. Production experiments following this paradigm have been shown to elicit responses that, when averaged over several speakers, replicate distributional facts about the lexicon (e.g. Zuraw 2000; Ernestus & Baayen 2003, and many others).

5.1.1 Participants

Participants were adult native speakers of Tgdaya Seediq (N=10; 7 female; ages 45-76). All speakers were paid 500NTD (around $17) for their time. Of the 10 participants, 7 notably had slight experience training to be Seediq language teachers. Consequently, they had metalingusitic awareness of suffixes and their functions, but were not taught explicitly about the vowel alternation processes.

5.1.2 Procedure

Since the experiment took place during the COVID-19 pandemic (August 2020), it was conducted remotely by the author, through video conferencing software. Stimuli were presented in Seediq orthography using Microsoft PowerPoint; each word was given in its own slide, and accompanied by its gloss in Chinese orthography.

The experimenter prompted speakers to give suffixed forms for stimuli by providing an existing paradigm (e.g. [hediq]∼[hudiqan]), and asking them to fill out the paradigm of the test item. This method worked well for the participants, as most have had some (limited) experience training to be Seediq language teachers, and therefore had metalinguistic knowledge about the suffixes.

If a participant failed to produce suffixed forms, the experimenter prompted them by providing more real stem-suffix examples. To minimize priming effects, example stem-suffix pairs always had /i/ as V2 (since the experiment included no stimuli where V2 was /i/). Subjects occasionally fluctuated between -an and -i suffixes, but this never resulted in variation of the stem allomorph. When needed, the experimenter would also provide a meaning for the stimulus verb. For example, the inflected form for *daruk* ‘oil, fat’ could have the meaning ‘to render the fat (out of food)’.

Prior to the experiment, speakers were asked to read a list of Seediq nouns to confirm their fluency with the orthography. Starting with two real-word practice items, speakers were asked to provide the /-an/ suffixed form for each word. After each item, the experimenter checked whether the speaker already knew the inflected forms of the stem (items known to speaker were excluded).

5.1.3 Stimuli

In a pilot experiment, speakers raised concerns that the use of nonce words in experiments would interfere with ongoing language revitalization efforts. In response to these concerns, the current study used ‘gapped forms’, or stems with no known suffixed forms, in place of nonce words. A full list of stimuli is given in the Appendix.

Gapped forms were selected using the following methods. First, most stimuli were formed by affixing noun stems with a verbalizer prefix ‘pu-’[9]. For example, the stimulus [puˈgakac] ‘VERB-chair’ can be interpreted as meaning ‘to build a chair.’ To sufficiently cover all experimental conditions, I also included some low-frequency, relatively unknown verbs.

I worked with a primary consultant (age 76, female) to confirm that all test stimuli had no known suffixed

[9] The prefix [pu-] can act as a “verbalizer” that derives a verb from a non-verb class Holmer (1996) describes this as a causative prefix, but I found that it could be used somewhat productively to form denominal verbs; the same verbalizer prefix is found in closely related languages like Squliq Atayal (Huang & Hayung 2008).

forms. A stem was determined to have no known suffixed form if she had never heard it before and had never heard her elders using it before. According to my consultant, Seediq speakers rarely use innovative suffixed forms, and therefore have a clear intuition of whether a stem is gapped. My primary consultant also has certification as a Seediq language teacher, so it was relatively straightforward to ask her whether specific suffixed forms existed.

Stems were determined to be plausibly suffixable by running a pilot experiment with two consultants; words they judged to be impossible to suffix were omitted from the final stimuli.

Stimuli consisted of disyllabic stems ending in closed syllables (i.e. CVCVC), where the first vowel (V1) was one of /a, e, u/ and the second vowel (V2) was one of /a, u/. This results in six possible vowel combinations, summarized in Table 3. These vowel combinations were selected to elicit a range of environments in which post-tonic [u] is expected to either alternate with [e] or not alternate. Stems with a post-tonic /a/ are expected to never show V2 alternations.

There were 8 test items for each vowel combination, as well as 24 filler items (4 per vowel combination), which were Seediq words with known suffixed forms. This resulted in a total of resulted in 72 stimuli ($8 \times 6 + 24$).[10]

5.2 Predictions

In the Seediq lexicon, the post-tonic [u] of a stem can potentially alternate with [e] or [o] in the suffixed form, and this alternation follows a vowel matching tendency. On the other hand, given a non-suffixed stem, post-tonic [a] and [i] never alternate.

Predicted speaker responses, based on these distributional generalizations, are summarized in Table 3; for each condition, an example stimulus is provided with the predicted preferred outcome given in parentheses. The rightmost column (‘Match’) indicates whether the predicted output results in vowel matching. If speakers generalize the vowel matching pattern, they should apply the [u]∼[e] alternation to most CeCuC stimuli. In CaCuC stems, [u]∼[e] alternation should be very infrequent. In CuCuC stems, vowel alternation should never be observed, since the faithful non-alternating outcome already satisfies vowel matching.

V1	V2	Prediction	Example	Match
a	u	disprefer alternation	ˈdaruk (duˈruk-an) ‘oil’	
e	u	[u]∼[e] alternation	ˈkeruŋ (kuˈreŋ-an) ‘wrinkles’	✓
u	u	never alternate	ˈcuguk (cuˈguk-an) ‘Bidens plant’	✓
a	a	V2 never alternates	ˈsabak (suˈbak-an) ‘dregs, pulp’	✓
e	a		ˈrehak (ruˈhak-an) ‘seed’	
u	a		kuˈsuwak (kusuˈwak-an) ‘yawn’	

Table 3: Experimental conditions: vowel alternations

5.3 Results

During the experiment, speakers always applied pretonic vowel neutralization. For example, given an input stimulus [ˈsabak], responses like [saˈbak-an] were never observed. In addition, speakers always applied stress shift; consistent with the lexicon, stress is exceptionlessly penultimate.

Experimental results for post-tonic vowel alternation are summarized in Fig. 3, which shows the proportion of response types by vowel condition, compared against lexical statistics. Cases where vowel alternation

[10] Items also varied by the identity of their final consonant, to test a set of consonant alternations. Results on final consonant alternation are not reported because they are not relevant to the current discussion on vowel matching.

obeyed the vowel matching pattern (i.e. resulted in the stem and suffixed forms having the same stressed vowel) are indicated in green.

Note that for 56 tokens (12% of total responses), speakers did not provide any responses. In a small subset of tokens (n=8, 2% of total responses), instead of inflecting the provided stem, speakers would provide the inflected form of an existing, segmentally similar verb. These tokens were omitted in Fig. 3.

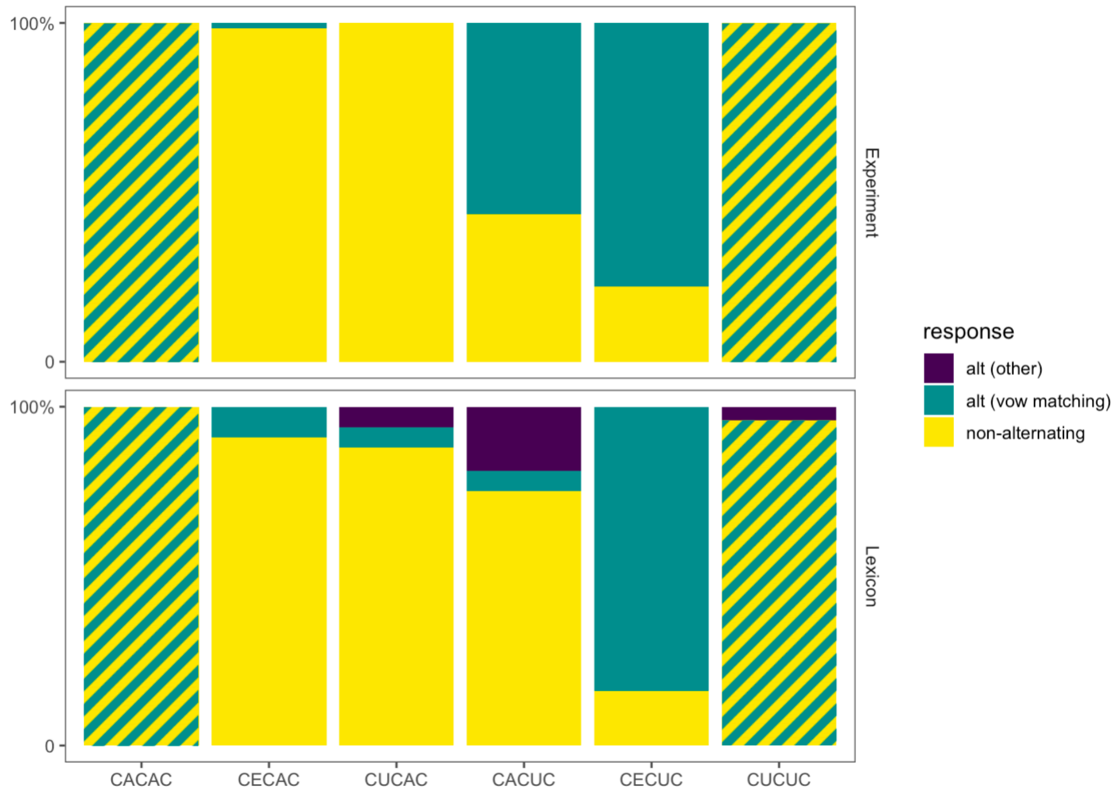

Figure 3: Vowel alternation rates in experiment vs. lexicon

First, looking at the left-hand columns, which shows results for stems with a post-tonic [a], we see that as expected, final /a/ almost never alternates. Note that for CaCaC forms (e.g. [ˈgakac∼guˈkac-an], the non-alternating form already satisfies vowel matching. There was one exceptional token (huˈrenaŋ∼huruˈneŋi), where an [a∼e] was observed; this alternation resulted in vowel matching. On the right-hand columns, CeCuC stems prefer the [u]∼[e] alternation, and CuCuC stems never alternate.[11] Again, note that for CuCuC forms, the non-alternating suffixed form spuriously satisfies vowel matching. In general, for both CeCuC and CuCuC stems, speakers' rates of alternation (or non-alternation) closely match the lexical statistics.

However, speakers deviated from the lexicon in CaCuC stems. [u]∼[e] alternation was not observed at all. Instead, for around half of the stems in this category, speakers applied a [u∼a] alternation (e.g. ˈdar**u**k∼duˈr**a**k-an). This alternation is irregular and novel, in the sense that it is not predicted by lexical statistics. Instead, it appears that speakers have extended the vowel matching pattern to CaCuC stems.

5.4 Discussion and interim summary

Overall, experimental results suggest that speakers have productively learned the vowel matching pattern. In fact, they also appear to have generalized this pattern beyond CeCuC and CoCuC stems, resulting in [u∼a]

[11] The lexicon contains one exceptional CuCuC form that alternates(ˈcuguk∼cuˈgak-an).

alternations for CaCuC stems.

As addressed in Section 1, Seediq vowel matching is reminiscent of harmony and copy epenthesis (Stanton & Zukoff 2018). Both harmony and copy epenthesis have been the subject of extensive analytical work, and can be treated as feature spreading in autosegmental accounts (e.g. Clements 1986; Kawahara 2007), or enforced by constraints on surface feature agreement (e.g. Hayes & Londe 2006). Copy epenthesis has also been analyzed using segmental correspondence, by enforcing faithfulness between a host segment and its copy (McCarthy & Prince 1995).

Notably however, approaches which enforce surface agreement do not straightforwardly capture Seediq vowel matching in OT. For example, when given a stimulus item such as [ˈpetus], speakers tended to provide the novel suffixed form [puˈtesan]. Because the pretonic vowel is always reduced, if we look only at the suffixed form [puˈtesan], there is no surface evidence of vowel matching. As a result, constraints such as Agree will not capture the tendency for vowels to agree.

Seediq vowel matching can only be captured by comparing morphologically related surface forms (specifically, stem and suffixed forms). Output-output correspondence constraints can do exactly this, by enforcing faithfulness between related output forms. However, classical correspondence constraints, which compare linearly related segmental units, cannot motivate vowel matching in Seediq. Instead, as already outlined in Section 2, I formalize Seediq vowel matching as the effect of prosodic correspondence between stem and suffixed allomorphs.

6 A stochastic OT analysis of vowel matching in Seediq

In this section, in order to confirm the claims about vowel copying made above, I flesh out an Optimality Theoretic analysis of vowel matching in Seediq. This analysis will account for the patterns of of vowel matching found in the experiment.

Vowel matching is gradient in Seediq, and did not apply categorically to novel forms in the experiment. To capture this, I assume the framework of Maximum Entropy (MaxEnt) Harmonic Grammar (Goldwater & Johnson 2003; Smolensky 1986), which is a stochastic variant of OT.

MaxEnt is a probabilistic variant of Harmonic Grammar (Legendre, Miyata & Smolensky 1990; Pater 2009), which are themselves variants of OT that use weighted (instead of ranked) constraints. MaxEnt generates a probability distribution over the set of candidate outputs based on their violations of a set of weighted constraints. Unlike classic OT, where strict ranking ensures that losing candidates never surface, all candidates in MaxEnt grammars receive some probability. However, if constraint weights are sufficiently different, MaxEnt produces results that are functionally very similar to classic OT, where the winning candidate gets near-perfect probability, while losing candidates get near-zero probability.

In all following tableaux, each constraint is associated with a weight; all constraint weights were learned using Excel Solver (Fylstra, Lasdon, Watson & Waren 1998), with the Conjugate Gradient Descent method. In each tableau, the column labeled **P** indicates the probability of a candidate occurring. For ease of interpretation, very small probabilities (on the order of 10^{-5}) are listed as zero.

The analysis laid out in this section is meant to demonstrate how Nuc-OO-Ident[F] is able to account for Seediq vowel matching. Because speaker's responses in the production experiment is assumed to better reflect how vowel matching applies in the synchronic grammar, the model will be fit to the experimental data, rather than the lexicon.

Section 6.1 introduces the segmental faithfulness constraints used in my analysis. Following this, I will focus on an analysis fit to experimental results. In Section 6.2, I briefly discuss pretonic vowel neutralizations, and constraints necessary for enforcing them. Section 6.3 outlines how prosodic correspondence constraints are used to capture the vowel matching pattern. Finally, in Section 6.5, I discuss the discrepancies between the lexicon and experimental results, and use Zuraw's (2000) dual listing/generation model to explain some of these differences.

6.1 Faithfulness constraints

In the current analysis, inputs to the model are assumed to not have listed suffixed forms, just like the experimental stimuli. As such, when deriving the suffixed form, the inputs are surface stem forms, and candidates are in an output-output (OO) correspondence relationship with input stems. Input-output (IO) faithfulness is not central to the current analysis, but their effects are discussed in Section 6.5.

The effect of traditional segmental OO-faithfulness is demonstrated in tableau (21), which shows how suffixed forms are derived for the hypothetical input stem [ˈpetus].[12] The constraint which enforces vowel matching will be introduced in Section 6.3; foe now, it is written as MatchV. Candidate (b), which undergoes vowel alternation to satisfy MatchV, violates segmental OO-faithfulness constraints, written in the tableau as OO-Faith. However, because OO-Faith has lower weight than the constraint which enforces vowel matching, candidate (b) is still preferred.

(21) *OO-faithfulness*

[ˈpetus]+/an/	MatchV	OO-Faith		
Weights	7	5	$\mathcal{H}$	*P*
a. puˈtusan	1		7	**0.12**
b. puˈtesan		1	5	**0.88**

6.2 Pretonic vowel reduction

Pretonically, vowels either delete, assimilate to a stressed vowel, or reduce to [u]. All three patterns can be motivated by fairly standard markedness constraints. In the interest of space, I discuss only pretonic vowel reduction.[13]

Pretonic vowel reduction to [u] is enforced by a positional licensing constraint (Crosswhite 2004). Specifically, I use License[u]/pretonic (Lic[u]/pret); this constraint is defined in (22), and essentially penalises non-[u] syllables in pretonic position.

Pretonic vowel reduction is exceptionless even in the experimental items, which have no listed suffixed form. This means that it must out-weigh (or outrank) all competing OO-Faithfulness constraints. This is demonstrated in tableau (23) for the hypothetical input [ˈpatas].

The winning candidate (b) violates Ident[high] due to alternation of pretonic /a/ with [u]. However, the faithful candidate (a) fatally violates a higher weighted Lic[u]/pret, and is therefore eliminated. Candidates which repair the markedness violation through other strategies are eliminated by higher weighted OO-faithfulness constraints. For example, candidate (c), which repairs the violation of Lic[u]/pret by deleting the pretonic vowel, is ruled out by the highly weighted constraint MaxC.

(22) License[u]/*pretonic*: In pretonic position, vowels must be [u] (i.e. non-[u] vowels cannot appear in pretonic syllables).

[12] Note that though it is not discussed here, both candidates (a) and (b) undergo pretonic reduction.
[13] Analyses for vowel assimilation and deletion can be found in Yang (1976).

(23) *Pretonic vowel reduction*.

[ˈpatas] +/an/	P	$\mathcal{H}$	*Lic([u], pret)* 20	*OO-ID[high]* 5	*MaxC* 10
a. paˈtas-an	0	20	1		
b. puˈtas-an	1	1		1	
c. ˈtas-an	0	10			1

Because pretonic vowel reduction is exceptionless and does not interact with the vowel matching alternation, subsequent tableaux will assume that Lic[u]/pret is never violated, and omit candidates such where the pretonic vowel is not reduced.

6.3 Vowel matching

In my analysis, the tendency towards vowel matching is formalized as the effect of an Ident constraint between the stressed nuclei of stem and suffixed forms. In Seediq, vowels are only fully constrastive in the stressed syllable, since vowels are reduced both pretonically and post-tonically. It is therefore plausible that the stressed nucleus would be given privileged status in the phonological representation. The constraint I adopt was defined above in Section 2, but is given again in (24) for the feature [high]. Nuc-Ident-OO[high] enforces identity of the [high] feature in the stressed nuclei of morphologically related output forms.

(24) Nuc-Ident-OO[high]: For α, a stressed nucleus of the base, and β, a stressed nucleus of the affixed form, $\alpha\mathfrak{R}\beta$, if α is [γhigh], then β is [γhigh].
i.e. Nuclei of corresponding stem/suffixed forms should agree in value for [high].

This correspondence constraint formulation is unusual in that is references prosodic units (stressed nuclei), but compares a *segmental* feature (vowel height) within these prosodic units. The comparison is not so unusual, however, if we consider that vowel quality is closely tied to prosodic structure. For example, vowel sonority (largely corresponding to height) constrains vowels from appearing as foot-peaks or foot-margins (Kenstowicz 1994). Vowel alternation also often references prosodic positions, with reduction happening in non-prominent positions, and contrast-enhancing alternations happening in prominent positions (Crosswhite 1999).

Ident constraints typically reference feature specifications, as in (24). For Seediq, the vowel matching pattern always results in complete vowel identity. As such, instead of referencing feature-specific Ident constraints, I will simply use the constraint **Nuc-Ident-OO[F]**, where **F** stands for a collection of faithfulness constraints that require two vowels to be identical.

The tableaux in (25) demonstrate how Nuc-OO-Ident[F] is used to derive vowel matching alternations for CVCuC forms. The column labeled ‘Obs’ shows the rates of alternation observed in the experimental stimuli. Stressed syllables are indicated in boldface for ease of interpretation.

For the representative input [ˈputuk], the faithful candidate (a) satisfies Nuc-OO-Ident[F], and therefore receives high probability (≈1). For [ˈpetuk], the high weight of Nuc-OO-Ident[F] relative to competing segmental OO-faithfulness constraints (OO-Ident[front] and OO-Ident[back]) causes candidate (d), which undergoes [u]∼[e] alternation, to be preferred over the faithful candidate.

For the input [ˈpatuk], candidate (f) is ruled out because it violates both segmental and prosodic OO-faithfulness. However, candidate (g), which undergoes post-tonic [u]∼[a] alternation to resolve Nuc-OO-Ident[F] violations, is slightly preferred over the faithful candidate (e). Compared to the input [ˈpetuk], the rate of vowel-matching for [ˈpatuk] is lower (0.56). This falls out from the relative weighting of segmental OO-faithfulness constraints. Specifically, candidate (g) violates OO-Ident[low], which has a relatively higher weight compared to the other segmental Ident constraints. Note that these tableaux also contains the constraint *P/i,u, which is not relevant to the current inputs, but will be explained in the following section.

(25) *Tableau: post-tonic vowel alternations for CVCuC inputs.* The probability of each candidate in the lexicon (Obs.) is shown alongside model predictions (P).

	Obs	P	H	*Nuc-OO-Ident[F]*	*OO-ID[low]*	*OO-ID[front]*	*OO-ID[back]*	**P/i,u*
				6.6	10.4	9.4	3.2	7.3
[ˈ**pu**tuk]+/an/								
a. puˈ**tu**kan	1	1	10.3					1
b. puˈ**te**kan	0	0	25.0	1		1	1	
[ˈ**pe**tuk]+/an/								
c. puˈ**tu**kan	0.22	0.22	18.4	1				1
d. puˈ**te**kan	0.78	0.78	17.0			1	1	
[ˈ**pa**tuk]+/an/								
e. puˈ**tu**kan	0.44	0.44	14.5	1				1
f. puˈ**te**kan	0	0	22.3	1		1	1	
g. puˈ**ta**kan	0.56	0.56	14.3		1		1	

Note that although it is not shown in the above tableaux, there should also be a constraint enforcing the correspondence relationship between the stressed nuclei of related output forms. These would have a similar function to Rose and Walker's (2004) Corr-CC constraints, which require segments of output strings to correspond. In the rest of this analysis, constraints enforcing correspondence are assumed to be highly weighted and essentially inviolable.

6.4 Explaining asymmetries in vowel alternation

In the experimental results, speakers applied a novel alternation to post-tonic [u], resulting in [u]∼[a] alternations. However, they never applied [a]∼[u] alternations to post-tonic [a], even when doing so would resolve violations of Nuc-OO-Ident[F].

There are multiple possible explanations for this; speakers could, for example, have learned a source-oriented generalization about the type of vowels allowed to undergo alternation Becker & Gouskova (2016). Alternation could also be explained in terms of anti-faithfulness constraints motivation alternation of post-tonic [u] (Alderete 2001).

I chose to explain this directionality as a preference for more sonorous vowels in stressed positions; [u]∼[a] alternations are preferred over [a]∼[u] alternations because the former increases the sonority of the stressed syllable, while the latter does the opposite. A similar preference for sonorous vowels in prosodic heads or foot-peaks (i.e. main-stressed syllables) has been observed in various languages, including Zabiče Slovene (Crosswhite 1999) and Chamorro (Chung 1983).

Kenstowicz (1994) formalizes this preference for sonorous vowels in foot-peaks using a family of constraints, where less sonorous vowels are relatively more constrained from appearing in stressed positions.

These constraints are given in (26); they penalize certain vowels in word-peaks, and are argued to follow a universal ranking hierarchy. In a weighted constraint model like MaxEnt, this means that a constraint like *P/ə should always have higher weight relative to *P/e,o.

(26) *Universal hierarchies and rankings for foot-peaks* (Kenstowicz 1994).
  a. Hierarchy for foot peaks: a > e,o > i,u >ə
  b. Constraint formulation: *P/*x*, assign a violation to every vowel *x* that is in a foot peak (i.e. the nucleus of a stressed syllable).
  c. Constraint ranking: *P/ə ≫ *P/i,u ≫ *P/e,o ≫ *P/a

I will adopt this approach, and specifically use the constraint *P/i,u. Other foot-peak constraints such as *P/a are assumed to be in the grammar, but are not active because of their relatively low weight.

The effect of *P/i,u is demonstrated in tableau (27), which compares inputs [ˈpatus] and [ˈputas]. Given the input [ˈpatus], candidate (b), which exhibits [u]∼[a] alternation, receives high weight because it satisfies both Nuc-OO-Ident[F] and *P/i,u. In contrast, for the input [ˈputas], the [a]∼[u] alternating candidate (c) receives low weight despite satisfying Nuc-OO-Ident[F]. This is because alternation increases violations of *P/i,u.

(27) *Tableau: effect of *P/i,u*

	Obs	P	H	*Nuc-OO-Ident[F]*	*OO-ID[low]*	*OO-ID[front]*	*OO-ID[back]*	**P/i,u*
				6.6	10.4	9.4	3.2	7.3
[ˈ**pa**tuk]+/an/								
a. puˈ**tu**kan	0.44	0.44	14.5	1				1
b. puˈ**ta**kan	0.56	0.56	14.3		1		1	
[ˈ**pu**tak]+/an/								
c. puˈ**tu**kan	0	0	19.7		1		1	1
d. puˈ**ta**kan	1	1	9.1	1				

This analysis makes predictions that should be tested in future work. Specifically, speakers are expected to disprefer extension of vowel matching when doing so reduces the stressed vowel's sonority. For example, the post-tonic vowel in CiCaC stem could undergo [a]∼[i] alternation to satisfy Nuc-OO-Ident[F]. However, because alternation of [a] to [i] increases violations of *P/i,u, [a]∼[i] alternation should occur at a lower rate than [u]∼[a] alternation did for CaCuC stems.

6.5 Lexicon vs. experiment

The model above was fit to experimental results. However, there are two discrepancies between lexical statistics and experimental results. In this section, I discuss these differences, and how they could potentially be accounted for in a model of Seediq vowel matching that aims to capture facts of both the lexicon and experiment.

The first discrepancy concerns lexical variation. Unlike the experimental stimuli, Seediq stems with known suffixed forms never show token variation in terms of whether they alternate, even if the resulting suffixed form violates vowel matching. The invariance of lexical items can be straightforwardly dealt with using Zuraw's (2000; 2010) DUAL LISTING/ GENERATION MODEL. Note that although I have adopted the lexical listing approach here, there are alternative solutions which also account for lexical specificity, such as constraint cloning (Becker 2009).

Under the DUAL LISTING/GENERATION MODEL (Zuraw 2000), both the stem and suffixed allomorphs of existing words are listed in the grammar. For example, given the stem-suffix pair [ˈremux]∼[ruˈmuxan], both /remux/ and /rumuxan/ are listed.

The grammar includes markedness constraints and output-output faithfulness constraints, whose relative weights give rise to speakers' gradient judgement of novel forms. However, highly weighted input-output faithfulness constraints protect listed forms from variation.[14]

For example, consider the tableau in (28), which shows a simplified derivation of the listed suffixed form [ruˈmuxan], which is associated with the stem [ˈremux]. The winning candidate (a) [ruˈmuxan] actually violates NUC-OO-IDENT[F]. However, we never observe candidate (b), [ruˈmexan], because /rumuxan/ is listed as a lexical entry. Highly weighted IO-faithfulness constraints, written as IO-FAITH on the tableau, rule out candidates such as (b). Note that candidate (b) also violates segmental OO-faithfulness, since the final vowel surfaces as [e] instead of [u].

(28) *Example of IO-faithfulness in lexically listed forms*

/rumuxan/∼[ˈremux]	IO-FAITH	NUC-OO-IDENT[F]	OO-FAITH		
Weights	20	7	5	$\mathcal{H}$	*P*
a. ruˈmuxan		1		7	**1**
b. ruˈmexan	1		1	25	**0**

In contrast to words with known suffixed forms, the experimental stimuli have no listed suffixed form, so vowel alternations will not result in violations of IO-FAITH. This is demonstrated below in (29). Because the input stem [ˈpetus] has no listed suffixed form UR, candidate (b), which undergoes vowel alternation to satisfy NUC-OO-IDENT[F], doesn't violate IO-FAITH, and gets assigned higher probability.

(29) *No IO-faithfulness when suffixed form is unlisted*

[ˈpetus]+/an/	IO-FAITH	NUC-OO-IDENT[F]	OO-FAITH		
Weights	20	7	5	$\mathcal{H}$	*P*
a. puˈtusan		1		7	**0.12**
b. puˈtesan			1	5	**0.88**

[14] Zuraw's (2010) model also employs USELISTED constraints, which require a listed entry to be employed. For the purposes of the present analysis, since stem-suffix pairs are not known to show token variation, USELISTED is assumed to be inviolable and highly weighted.

The second difference between the lexicon and experiment is more surprising; as discussed in Section 5, speakers in the experiment actually learned the vowel-matching pattern non-veridically. Specifically, speakers over-generalized vowel matching, and applied a novel [u]∼[a] alternation to CeCuC stems. In contrast, the lexicon has only [u]∼[e] and [u]∼[o] alternations (barring a few exceptions).

There are various possible reasons for why speakers may have preferentially over-generalized vowel matching. Over-extension of vowel matching could have been due to a simplicity bias, causing speakers to learn a general vowel matching constraint, rather than a more complex constraint that is specific to only mid vowels. This analysis would be in line with a body of work suggesting that people preferentially learn simpler constraints (e.g. Moreton & Pater 2012). On the other hand, speakers may have preferentially learned vowel matching because it is somehow less marked, and guided by UG principles. This type of naturalness bias has been explored in various work, including Becker et al. (2011) and Hayes, Spitar, Zuraw & Londe (2009).

Regardless of the exact reason, an ideal model of Seediq vowel matching should be able to capture speakers' non-veridical learning of vowel matching. This type of learning model would take the lexical data as input, and be able to predict the experimental results, where Seediq learners over-extended vowel matching. Developing this model is beyond the scope of the current paper, and should be addressed in future work. In particular, MaxEnt models which implement a bias, such as the ones laid out in Wilson (2006) and White (2017), are promising, and could potentially be adapted to the Seediq data.

7 Could Seediq vowel matching have a diachronic explanation?

The post-tonic u-e alternation in Tgdaya Seediq results from a sound change of Proto-Austronesian (PAn) *ə to [u] in the final syllable, and to [e] in other environments (Li 1981). As such, we should consider the possibility that modern Seediq's tendency towards vowel matching is an artifact of historical distributions. In other words, Seediq might historically have had much more CəCəC forms (than CəCuC/CəCiC/CəCaC forms), resulting in the current tendency towards vowel matching.

It is difficult to tell how much vowel matching is directly inherited from proto-Atayalic, which encompasses both Seediq and Atayalic dialects. This is because extensive vowel reduction has taken place in all dialects of proto-Atayalic. In particular, Proto-Austronesian (PAn) schwa has reduced to *u in the final position of *all* languages in proto-Atayalic (Li 1981). However, regardless of the exact historical origins of vowel matching, the results of the productivity test suggest that it is present in the synchronic grammar of Seediq speakers. In particular, Seediq speakers applied a [u]∼[a] alternation, which has no historic basis.

In addition, recent quantitative work on harmony in Oceanic languages by Alderete & Finley (2016) suggests that vowel matching is present as a gradient preference in Oceanic, and likely in Proto-Austronesian. However, this preference is stronger for the high and mid vowels, and does not apply to the low vowel /a/, which has a less restricted distribution and more freely co-occurs with all over vowels. In contrast, what I find is that speakers have extended vowel matching to /a/ (giving rise to alternations like [ˈpatus]∼[puˈtasan]). This again supports the conclusion that synchronically, Seediq speakers have learned a vowel matching pattern which cannot be explained by historical facts alone.

8 Conclusion

Based on a survey of 340 Seediq verb paradigms, the current study finds that Seediq paradigms show a tendency towards vowel matching, where stressed vowels of morphologically related forms match each other. This vowel matching pattern was shown to be productive, and even overgeneralized, in an experiment where speakers were given stems with no known suffixed forms, and instructed to provide novel suffixed forms.

The Seediq data provides novel evidence for PROSODIC CORRESPONDENCE as defined by Crosswhite (1998). Whereas Crosswhite uses the correspondence between stressed nuclei to explain a length alternation, I extend it to explain vowel alternations.

Prominent prosodic positions are known to be privileged with respect to standard segmental faithfulness (Kenstowicz 1995; Beckman 1999, etc.). As such, it is somewhat surprising that there are so few languages documented as having the type of prosodic correspondence found in Chamorro and Seediq, where prosodic positions correspond even when they are not linearly related. There are two possible reasons for this.

First, cases of prosodic correspondence could have been overlooked because they only emerge when we compare related *output* forms. In frameworks which focus on deriving surface forms from URs, such as generative phonology, cases of prosodic correspondence would be less evident. One potential case of this variety comes from Maga Rukai (Austronesian), which has a vowel lowering process that cannot be captured by surface agreement constraints. This alternation has been analyzed in generative rule-based phonology as height assimilation/vowel coalescence followed by syncope (Hsin 2000). However, it could potentially benefit from a prosodic correspondence account, as the result of OO-correspondence between specific syllabic positions of the [high] feature.

Another potential reason for the lack of evidence for prosodic correspondence is that some of these cases are gradient, and only emerge when looking at statistical patterning in the lexicon. The approach taken in the current paper, where vowel matching was found using a combination of lexical statistics and experimental evidence, could prove useful in finding non-categorical cases of prosodic correspondence.

Appendix A

Table 4 shows the full results of the logistic regression model reported in Section 4.2.

Model: VOWEL MATCH $\sim$ V1 + V2

	VOWEL MATCH		
Predictors	*Odds rations*	*CI*	*p*
(Intercept)	0.79	0.47-1.33	0.376
V1 [e]	0.75	0.41 - 1.37	0.351
V1 [i]	0.40	0.17 – 0.91	0.033*
V1 [o]	1.16	0.17 – 9.46	0.880
V1 [u]	0.96	0.49 – 1.89	0.908
V2 [i]	0.45	0.22 – 0.91	0.029*
V2 [u]	**2.34**	**1.39 – 3.96**	**0.001****

Table 4: Effect of V1 and V2 on vowel matching

Appendix B: stimuli

Test items			
V1	V2	word	gloss
a	a	papak	foot
		sabak	pulp (of fruit)
		kkarang	walk on all fours
		dayaN	make firebreak
		tanah	red
		tapaq	pat, slap
		slmadac/hlmadac	hunting knife
		gakac	chair

a	u	tatuk	xylophone
		daruk	fat, oil
		rapung	mold
		halung	gun
		damux	rooftop
		aguh	call over
		ahuc (paahuc)	hoe
		lapuc	lint
e	a	rehak	seed
		tpetak	conflict
		hrenang	sound
		gelang	string of, bundle of
		qseyaq	cough
		phepah	flower
		ngerac	outside
		sepac	four
e	u	etuk	contain
		thbehuk	stuffy
		kerung	wrinkles
		bngebung	grill
		gebuh	granules
		knedux	thick
		deluc	attached to
		peeruc	pillar
u	a	ksuwak	yawn
		ptkurak	month-old celebration for infant
		kurang	gums
		rubang	hunting tool
		hrulas	spit
		srmusaq	turbid
		hunac	point
		murac	compress
u	u	cuguk	Bidens plant
		kduruk	forehead
		bukung	hunch-backed
		ubung	loom
		btunux	rock
		pnunuh	breastfeed
		gukuc	wheel
		hukuc	cane
Filler items			
V1	V2	word	gloss
a	a	awak	lead (by leash)
		qbahaN	listen
		qamas	pickle
		qaras	happy
a	u	saruk	burn (fur, hair)
		balung	egg
		rahuq	leak away
		haNuc	cook, boil
e	a	pkepak	touch, feel
		gedaN	lose

		lepax	grind
		bcebac	cut
e	u	eluk	close
		geruN	split, break
		remux	enter
		keruc	(cut with) saw
u	a	sdurak	chase
		duraN	rope hunting trap
		luqah	injure
		squwaq	noisy
u	u	suyuk	twist into thread
		putuN	light (fire)
		plukus	wear (clothes)
		lutuc	ancestry, descendants

References

Alderete, John & Sara Finley. 2016. Gradient vowel harmony in oceanic. *Language and Linguistics* 17(6). 769–796. doi:10.1177/1606822X16660960.

Alderete, John D. 2001. Dominance effects as transderivational anti-faithfulness. *Phonology* 18(2). 201–253. doi:10.1017/S0952675701004067.

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

Becker, Michael & Maria Gouskova. 2016. Source-oriented generalizations as grammar inference in Russian vowel deletion. *Linguistic Inquiry* 47(3). 391–425. doi:10.1162/LING_a_00217.

Becker, Michael, Nihan Ketrez & Andrew Nevins. 2011. The surfeit of the stimulus: analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language (Baltimore)* 87(1). 84–125. doi:10.1353/lan.2011.0016.

Beckman, Jill N. 1999. *Positional faithfulness : an optionality theoretic treatment of phonological asymmetries*. New York: Garland.

Benua, Laura. 1995. Identity effects in morphological truncation. In Suzanne Urbanczyk Jill N. Beckman & Suzanne Urbanczyk (eds.), *University of massachusetts occasional papers 18: Papers in optimality theory*, 77–136. Amherst: GLSA.

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14(2-3). 150–177. doi:10.1080/00437956.1958.11659661.

Chung, Sandra. 1983. Transderivational relationships in Chamorro phonology. *Language* 59(1). 35–66. doi:10.2307/414060.

Clements, G.N. 1986. Syllabification and Epenthesis in the Barra Dialect of Gaelic. In Koen Bogers, Harry van der Hulst & Maarten Mous (eds.), *The phonological representation of suprasegmentals*, 317–336. Dordrecht: Foris. doi:10.1515/9783110866292-017.

Council of Indigenous People. 2020. Aboriginal household registration statistics data analysis in oct 2020. https://www.cip.gov.tw/portal/docDetail.html?CID=940F9579765AC6A0.

Council of Indigenous Peoples. 2020. Online dictionary of aboriginal languages. http://e-dictionary.apc.gov.tw/Index.htm. Accessed: 2020-09-30.

Crosswhite, Katherine. 1998. Segmental vs. prosodic correspondence in Chamorro. *Phonology* 15(3). 281–316. doi:10.1017/S0952675799003619.

Crosswhite, Katherine. 1999. *Vowel reduction in optimality theory.*: University of California, Los Angeles dissertation.

Crosswhite, Katherine. 2004. Vowel reduction. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 191–231. Cambridge University Press.

Ernestus, Mirjam & R Harald Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79(1). 5–38. doi:10.1353/lan.2003.0076.

Frisch, Stefan A, Janet B Pierrehumbert & Michael B Broe. 2004. Similarity avoidance and the ocp. *Natural Language & Linguistic Theory* 22(1). 179–228. doi:10.1023/B:NALA.0000005557.78535.3c.

Fylstra, Daniel, Leon Lasdon, John Watson & Allan Waren. 1998. Design and use of the Microsoft Excel Solver. *Interfaces* 28(5). 29–55. doi:10.1287/inte.28.5.29.

Goldwater, Sharon & Mark Johnson. 2003. Learning ot constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson & Osten Dahl (eds.), *Proceedings of the stockholm workshop on variation within optimality theory*, 111–120. Stockholm: Stockholm University, Department of Linguistics.

Hayes, Bruce & Zsuzsa Cziráky Londe. 2006. Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology* 23(1). 59–104. doi:10.1017/S0952675706000765.

Hayes, Bruce, Péter Spitar, Kie Zuraw & Zsuzsa Londe. 2009. Natural and unnatural constraints in hungarian vowel harmony. *Language (Baltimore)* 85(4). 822–863. doi:10.1353/lan.0.0169.

Holmer, Arthur. 1996. *A parametric grammar of seediq*: Lund University dissertation.

Hsin, Tien-Hsin. 2000. *Aspects of maga rukai phonology*: University of Connecticut dissertation.

Huang, Lillian M & Tali' Hayung. 2008. Syntax and semantics of p- in Squliq Atayal. *Language and Linguistics* 9(3). 491–521.

Kawahara, Shigeto. 2007. Copying and spreading in phonological theory: Evidence from echo epenthesis. In Ehren Reilly Leah Bateman, Michael O'Keefe & Adam Werle (eds.), *University of massachusetts occasional papers 32: Papers in optimality theory iii*, 111–143. Amherst: GLSA.

Kenstowicz, Michael. 1994. Sonority-driven stress. Ms, MIT. Available as ROA-33 from the Rutgers Optimality Archive.

Kenstowicz, Michael. 1995. *Base-identity and uniform exponence: alternatives to cyclicity*: MIT MA thesis.

Kiparsky, PAUL. 2000. Opacity and cyclicity. *Linguistic review* 17(2-4). 351–366. doi:10.1353/lan.0.0169.

Kiparsky, Paul. 2015. Stratal ot: A synopsis and faqs. In Yuchau E. Hsiao & Lian-Hee Wee (eds.), *Capturing phonological shades within and across languages*, 2–44. Cambridge Scholars.

Kitto, Catherine & Paul De Lacy. 1999. Correspondence and epenthetic quality. *Austronesian Formal Linguistics Association (AFLA) 4* .

Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky. 1990. Harmonic grammar—a formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. In *Proceedings of the 12th annual conference of the cognitive science society*, 884–891. Hillsdale, NJ: Lawrence Erlbaum.

Li, Paul Jen-kuei. 1991. Vowel deletion and vowel assimilation in Sediq. In Robert Blust (ed.), *Currents in pacific linguistics: Papers on austronesian languages and ethnolinguistics in honour of George W. Grace*, vol. 117 of Pacific Linguistics Series C, 163–169. Canberra: Australian National University.

Li, Paul Jen-kui. 1981. Reconstruction of Proto-Atayalic phonology. *Bulletin of the Institute of History and Philology* 52.

McCarthy, John. 1995. Faithfulness in prosodic morphology and phonology: Rotuman revisited. Ms, University of Massachusetts, Amherst.

McCarthy, John J. & Alan S. Prince. 1995. Faithfulness and Reduplicative Identity. In Laura Walsh Dickey Jill N. Beckman & Suzanne Urbanczyk (eds.), *Papers in optimality theory*, 249–384. Amherst: GLSA.

Moreton, Elliott & Joe Pater. 2012. Structure and substance in artificial-phonology learning, part i: Structure. *Language and linguistics compass* 6(11). 686–701. doi:10.1002/lnc3.363.

Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive science* 33(6). 999–1035. doi:10.1111/j.1551-6709.2009.01047.x.

Prince, Alan & Paul Smolensky. 1993, 2004. *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell.

Rose, Sharon & Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language (Baltimore)* 80(3). 475–531. doi:10.1353/lan.2004.0144.

Smolensky, Paul. 1986. Information processing in dynamical systems: Foundations of harmony theory. Tech. rep. Colorado Univ at Boulder Dept of Computer Science.

Stanton, Juliet & Sam Zukoff. 2018. Prosodic identity in copy epenthesis. *Natural Language & Linguistic Theory* 36(2). 637–684. doi:10.1007/s11049-017-9385-9.

Steriade, Donca. 2000. Paradigm uniformity and the phonetics-phonology boundary. In Janet Pierrehumbert & Michael van Broe (eds.), *Papers in laboratory phonology v: acquisition and the lexicon*, 313–334. Cambridge: Cambridge University Press.

White, James. 2017. Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a p-map bias. *Language* 93(1). 1–36. doi:10.1353/lan.2017.0001.

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive science* 30(5). 945–982. doi:10.1207/s15516709cog0000_89.

Yang, Hsiu-fang. 1976. The phonological structure of the paran dialect of Sediq. *Bulletin of the Institute of History*

and Philology Academia Sinica 47(4). 611–706.
Zuraw, Kie. 2000. *Patterned Exceptions in Phonology*: University of California, Los Angeles dissertation.
Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to tagalog nasal substitution. *Natural Language & Linguistic Theory* 28(2). 417–472. doi:10.1007/s11049-010-9095-z.

Author name
Department of Linguistics
Affiliation
Address1
Address2
email.edu