

Jonathan David Bobaljik
Disharmony and decay

Itelmen vowel harmony in the 20th century

1 Introduction

The Chukotko-Kamchatkan (CK) languages have a characteristic dominant-recessive, bi-directional vowel-harmony process, in which both roots and affixes alternate: an underlying recessive vowel alternates with its dominant counterpart, if there is a dominant element elsewhere in the word. Itelmen [itl], whose status in the CK group is still debated, shows vowel alternations that appear to be the remnants of the standard CK harmony system, such as in (1):

- (1) a. **isx** **esx**-anke
 ‘father’ ‘father-DAT’
 b. k-**siŋ**-qzu-kne’n **seŋ**-zo-z-in
 ‘PRT-fly-ASP-PRT.PL’ ‘fly-NDIR-PRES-3SG’

I have two goals in this paper. First, I will attempt to establish that a harmony system essentially of the CK sort was robustly active in Itelmen as recently as a century ago, drawing on an analysis of Itelmen texts collected at the beginning of the 20th century. In this, I concur with other authors who argue that the harmony system was all but lost over approximately three generations, remaining in vestigial form as a morphologized ablaut system affecting some (but by no means all) roots and affixes (Volodin, 1976; Asinovskij & Volodin, 1987).¹

The second goal is to ask why the harmony system collapsed in such a short span of time. In keeping with current events, I blame Russian influence. Specifically, I suggest that the sudden and drastic decline of harmony was a result of language contact – an influx of disharmonic loanwords, primarily from Russian (although also possibly from neighbouring Koryak *a*-dialects). More narrowly, I

¹ The system attested in 1910 was very similar to the general CK pattern, but is not identical, notably, as Volodin stresses, in the apparent absence of *e*-*a* alternations in verbal roots. The question of the origin of the system is therefore related to the question of whether Itelmen is a divergent member of the CK family, or a distinct language whose many shared grammatical properties are the result of extended language contact (the latter is the view held by Volodin). While I tend towards an account of the similarities in terms of a common ancestor, the focus of this paper is an understanding of the change in Itelmen across the 20th century, for which the question of the original (pre-1910) nature and source of the Itelmen harmony system does not need to be resolved.

report here on a preliminary attempt to characterize this effect quantitatively, comparing texts collected by W. Jochelson in 1910–1911 (Worth, 1961) to texts collected in 1993–1994, and making use of the general approach to rule learnability proposed in Yang (2016). The leading idea is as follows: the learner of Itelmen (or of any CK language) is faced with vowel alternations in some, but not all morphemes, and must decide whether these alternations are governed by a general phonological rule, or are simply listed variants (equivalently, the output of ‘minor’ phonological rules) that must be learned as such and listed as lexically-specific alternations. For Itelmen, any rule that is to be posited will have some number of exceptions. Yang’s approach offers a formula for calculating the robustness of any given phonological rule, a measure of the proportion of exceptions in a corpus with some quantifiable number of opportunities for the rule to apply. Yang posits a specific threshold which constitutes the number of exceptions to a given rule that can be tolerated while still maintaining a productive grammatical rule. With some (I hope plausible) assumptions about how to calculate this for Itelmen vowel harmony, I demonstrate here over a preliminary sample that the number of harmony exceptions falls below Yang’s Threshold in 1910, but vastly exceeds the threshold by 1994. Over that time period, the number of Russian loans in Itelmen texts increased substantially, plausibly to a sufficient extent to have rendered the harmony rule unlearnable.

Before discussing the quantitative evidence, I first introduce the standard Chukotkan harmony pattern. I then provide a close analysis of a sample of the texts collected by W. Jochelson in 1910–1911. While there is clear evidence of vowel alternations in these texts, there are also significant numbers of apparently disharmonic forms. When examined more carefully, many of the apparent exceptions in the old texts are best seen as artifacts of the transcription used in that source. Once idiosyncracies of transcription are controlled for, the vowel harmony rule turns out to be learnable, under Yang’s formula, in the 1910 texts. Applying the same criteria to a text collected in the 1990s, we find that the vowel harmony system had become unlearnable by that time, consistent with the main hypothesis that Itelmen has indeed lost (productive) vowel harmony.

2 Vowel harmony in Chukotko-Kamchatkan

Itelmen constitutes the Kamchatkan branch of the Chukotko-Kamchatkan family. Three hundred years ago, at the time of first European contact, there were (at least) three distinct Itelmen languages, together spoken by somewhere between 10–25,000 speakers across the southern half of the Kamchatka peninsula (Volodin, 2003, 27). By the early 20th century, the Itelmen-speaking population had been

decimated, with only the Western Itelmen language spoken in two dialect groups covering eight villages on the Okhotsk coast. As of 2017, the language is spoken natively and fluidly by fewer than five elderly members of the community. The Chukotkan branch of the family includes Chukchi, Koryak, Kerek and Alutor, the former two comprising multiple dialects and having a few thousand native speakers.

Of interest in this paper is the vowel harmony system, shared in one form or another by most of the modern languages and therefore generally reconstructed for the proto-language (Muravyova, 1979; Fortescue, 2005). The system is a dominant-recessive system. The Proto-Chukotkan vowel inventory is reconstructed as in (2), see (Muravyova, 1979; Fortescue, 2005), with some variation among authors regarding specifics of vowel quality.

(2) Proto-Chukotkan vowel inventory

recessive	i	u	ɛ
dominant	e	o	a
transparent	ə		

Non-schwa vowels in a morpheme, and indeed in a word, will be either all dominant, or all recessive (schwa, whether underlying or epenthetic, may occur with either set). Dominant vowels undergo no alternations, but recessive vowels change to their corresponding dominant counterparts if any morpheme in the same word has dominant vowels. Kenstowicz (1979) suggests that this may be an ATR harmony system, that is spreading of [-ATR] (i.e., root retraction), although various questions (especially of phonetic detail) are left unresolved (see in particular Krause, 1979; Calabrese, 1988 for discussion). I take no stand here on the actual feature involved and simply use a diacritic [$\pm D$] as a stand-in for whatever phonological feature turns out to be accurate. Thus, the following loose characterization of the vowel harmony rule will suffice for the purposes of this paper:

$$(3) \quad V^{[-D]} \rightarrow V^{[+D]} / [\{ \dots _ , (V)^{[+D]} \dots \}]_\omega$$

2.1 Vowel harmony in Chukchi: the ideal pattern

The Chukchi vowel harmony system, well discussed in the literature (Bogoras, 1922; Skorik, 1961; Krause, 1979; Kenstowicz, 1979; Calabrese, 1988; Dunn, 1999), presents a fairly conservative example which thus serves as a convenient point of departure to illustrate (3).

The Chukchi vowel inventory is given in (4). The vowels are divided into three pairs, each pair having a recessive and dominant member. The ambivalent status of *e* will be discussed shortly. In addition, there is a schwa, which is normally (but not always, see below) neutral and transparent to vowel harmony, undergoing no change in harmony contexts but not blocking the application of harmony across it.

(4) Chukchi vowel inventory

recessive	i	u	e ₁
dominant	e ₂	o	a
transparent	ə		

Other than schwa, a given Chukchi word normally contains only dominant or only recessive vowels. A dominant vowel in any morpheme causes all recessive vowels elsewhere in the word to be replaced by their dominant counterparts, as dictated by (3).

The examples in (5) show the alternation in affix vowels, controlled by the root. Affixes with a recessive vowel surface as such with recessive vowels, but the dominant alternants are used with roots containing dominant vowels.

(5) Root controls affix (prefix and suffix)

-(n)u DESIG	recessive:	/milute/ ‘rabbit’	milute-nu
		/tutlik/ ‘snipe’	tutlik-u
	dominant:	/wopqa/ ‘moose’	wopqa-no
		/orw/ ‘sled’	orw-o
(y(e))-...-(t)e INSTR	recessive:	/milute/ ‘rabbit’	ye-milute-te
		/kupre/ ‘net’	ye-kupre-te
	dominant:	/wala/ ‘knife’	wala-ta
		/rərka/ ‘knife’	ya-rərka-ta

The inverse pattern is shown in (6). Here, the roots alternate, surfacing with dominant vowels when the affix contains a dominant vowel, and with recessive vowels otherwise.²

² It has been claimed that there are quite generally no dominant prefixes in vowel harmony systems (see e.g., Baković, 2000, 228; see Moskal, 2015 for exceptions and discussion). Bogoras (1922) and (Skorik, 1977, 325) both claim that there are dominant prefixes in Chukchi. However I suspect that the items they identify are independent roots, suggesting an analysis of these forms as root-root compounds, rather than prefixes. The comitative in (6) is (descriptively)

(6) Affix controls root	ROOT	ABS	COMITATIVE /ɣ(a)-...-ma/
	/milute/ ‘rabbit’	milute-t	ɣa- melota -ma
	/titi/ ‘needle’	titi-ŋə	ɣa- tete -ma
	/rʔew/ ‘whale’	rʔew	ɣa- rʔaw -ma
	/ləle/ ‘eye’	ləle-t	ɣa- ləla -ma

Showing a typologically rare pattern, compounding (incorporation) structures show vowel harmony applying in combinations of two roots, with a dominant vowel in one root triggering lowering of a recessive vowel in another, regardless of direction (7):

(7) Root-Root interaction (incorporation)

ROOT	PREDICATE FORM	INCORPORATED	ROOT 2 GLOSS
/teŋ/ ‘good’	nə-teŋ-qin	taŋ -kawkaw	/kawkaw/ ‘zwieback’
		taŋ -čotčot	/čotčot/ ‘pillow’
/om/ ‘warm’	n-om-qen	om- penpen	/pinpin/ ‘ash’

2.2 Morphologization of harmony

Although the discussion above represents the basic Chukotkan system, there are a few respects in which the actual surface system of Chukchi deviates from the idealized system just outlined.

First, as noted in (4), there are two phonologically distinct /e/ vowels, one dominant, the other recessive. Recessive /e/ undergoes harmony and becomes [a], as shown in (8a), where the trigger is the comitative circumfix seen in (6). By contrast, the word for ‘road’ has a dominant /e/ which undergoes no alternation but itself triggers the alternation, in this case forcing an alternation on the incorporated adjective (just as in (7) above):

- (8) a. /rʔew/ ‘whale’ → ɣa-**rʔaw**-ma COMITATIVE
 b. /rʔet/ ‘road’ → **taŋ**-rʔet ‘good road’

circumfixal, and appears to contain a dominant vowel in the prefixal portion. The prefixal portion can instead be analyzed as the same (recessive) element as in the instrumental ɣ(e)-...-(t)e in (5). The trigger for harmony on this view would be the dominant vowel in the suffix, -ma.

While distinct in terms of their phonological behaviour, there are conflicting views in the literature as to whether dominant and recessive /e/ are phonetically distinct. Bogoras (1922), Skorik (1961, 22ff), and Asinovskij & Volodin (1987) report that the two /e/ vowels are distinct, while Mel'nikov (1948, 209), Fortescue (1998, 128), Dunn (1999) dispute this; for example, Dunn (1999, 48) states unequivocally: “there is no phonetic difference between” dominant and recessive [e].³ It may well be the case that the system requires diacritic marking of two distinct underlying /e/ vowels, as in (4).

A diacritic is needed in Chukchi in any event (and indeed in all Chukotko-Kamchatkan languages with vowel harmony) to account for morphemes that contain no full vowels and have only schwa (whether epenthetic or underlying) or no vowel at all. Some of these trigger harmony as if they contained a dominant vowel, while others do not (Krause, 1979, 13–14; Muravyova, 1979, 138–141). For example, the affixes in (9) have only schwa or no vowel at all, but trigger harmony alternations in the roots they attach to:

(9)	AFFIX	ROOT	SUFFIXED FORM
	a. -ytə	/milute/	melota -ytə ‘to the rabbit’
	b. -jpə	/titi/	tete -jpə ‘from the needle’
	c. -tk-	/utt/	ott -ə-tk-ən ‘crown of a tree’
	d. -lɣən	/milute/	melota -l-ɣ-ən ‘rabbit (singulative)’

Further contrasting pairs with schwa-containing roots are given in (10). In each pair, the first member fails to trigger harmony on a recessive suffix, while the second member triggers harmony on the same suffix.

(10) a.	GLOSS	ROOT	INFINITIVE
	i. sleep	/jəlq/	jəlq- et -ək
	ii. dark	/pəlm/	pəlm- at -ək
	b. GLOSS	ROOT	ADJECTIVE
	i. old	/ənpə/	n-ənpə- qin
	ii. dark	/pəlm/	nə-pəlm- qen

³ A related question is whether the vowels that are the output of the harmony rule are phonetically the same as underlyingly dominant vowels. Here too, reports diverge, with Skorik (1961) claiming that the derived dominant vowels are phonetically distinct from underlyingly dominant ones, and Bogoras (1922) and Dunn (1999) disagreeing. Calabrese (1988) sees the conflicting descriptions as evidence of dialect differences. The issue is clearly relevant to the proper understanding of the rule in (3). See also Kenstowicz (1979) and Krause (1979).

Examples of purely consonantal roots (i.e., roots with no underlying vowel at all) are given in (11). Those in group a. fail to trigger harmony, while those in group b. trigger vowel harmony, illustrated with the preterite circumfix.

(11)		ROOT	PRETERITE	GLOSS
	a.	i. /ŋt/	ye-nt-ə-lin	'he has cut off'
		ii. /rʏ/	ye-rʏ-ə-lin	'he has dug, scratched'
	b.	i. /tm/	ya-nm-ə-len	'he has killed'
		ii. /tw/	ya-tw-ə-len	'he has said'
		iii. /rw/	ya-rw-ə-len	'he has split'

Calabrese (1988) proposes that Chukchi may treat [-ATR] (i.e., [+D]) as a (diacritic) property of morphemes and not of vowels as such, interpreting (3) as applying at the morpheme tier, rather than the vowel tier. On this view, Chukchi would have only three full vowels underlyingly (rather than the three pairs in (4)), which undergo alternations according to whether they occur in a [+ATR] or [-ATR] environment. The morphological diacritic approach, of course, deals with the facts in (9)–(11) quite readily (compare the root markers of Lightner, 1965 and related work).

Whether we adopt such a proposal or not, what is important is that the evidence in (9)–(11) shows that diacritic marking is needed – the vowel harmony system is not entirely phonologically transparent. Nevertheless, it is internally consistent: the diacritic value of a morpheme cannot conflict with the value of the vowels in a word, yielding disharmonic words. Morphemes containing {*i,u*} are unambiguously recessive and will always undergo harmony alternations. Morphemes containing {*a,o*} are unambiguously dominant and will trigger harmony alternations on recessive morphemes. Morphemes containing only *e* and/or schwa or no vowels at all are ambiguous, and require diacritic marking (though the former may in fact be phonetically distinguished in some dialects).⁴

⁴ There are two instances where apparent surface disharmony is tolerated, both noted in Krause (1979). In the vocative only, a stressed schwa may be pronounced [o], with this pronunciation having no effect on harmony, see (i). There is also an optional rule rounding schwa to [u] before /w/. This segment may occur in a dominant environment and does not undergo a further change to [o]; see (ii). The existence of these forms is not widely commented on in the literature, but is potentially relevant to the question of how tolerant a productive rule may be to surface exceptions, taken up later in this paper.

i. $\hat{\text{a}} \rightarrow \hat{\text{o}}$: túmʏ-ət 'friend-PL' vs. tumʏ-ót 'O friends!' (Krause, 1979, 59)

ii. $\text{ə} \rightarrow \text{u}$ / _w : ətləwʏjot ~ ətluwʏjot 'grandchildren' (Krause, 1979, 116)

2.3 Postscript: the rest of Chukotkan

South of Chukchi, the vowel harmony patterns are less pristine, a fact that is relevant since Itelmen has been in extended contact with Koryak, rather than Chukchi. In some Koryak and Alutor dialects, recessive *e* (<**ε*) and (dominant) *a* have partially merged (Stebnickij, 1934; Muravyova, 1979, cf. Bogoras, 1917, 1922). This has led to a curious two-part harmony system, as described by Muravyova (1979); Abramovitz (2015): merged *a* may be described as ‘weakly dominant’; it does not affect recessive vowels {*i,u*}, allowing for words that have both *a* and *i* on the surface. However, weakly dominant *a* does trigger a limited harmony rule changing recessive *e* to *a*. The examples in (12) illustrate: *a*-dialect *kali* ‘write’ corresponds to Chukchi *keli*. In the (a) example, /*e*/ lowers to [a] under the influence of the weakly dominant /*a*/ in *kali*, but the /*i*/ remains unaffected. Addition of a truly dominant morpheme, such as *-jo* in (12b) lowers the /*i*/ to [e]. More curiously, just as a morpheme with no dominant vowels may be diacritically marked as dominant, so too may morphemes with no dominant vowels be diacritically specified as weakly dominant, as in (12c) from (Muravyova, 1979, 148), and confirmed by Rafael Abramovitz (pc and Abramovitz, 2015):

- (12) a. /*kali*/ ‘write’ + *-te* *kali-ta*
 b. /*kali*/ ‘write’ + *-jo* + *-te* *kale-jo-ta*
 c. /*quqlu*/ ‘make a hole’ + *ʔe*-...-*lin* *ya-quqlu-lin*

Muravyova (1979) notes that some dialects with the *e/a*-merger have gone even further. According to her description, Vyvenka Alutor has seen a merger of all dominant-recessive pairs, and has been reorganized as having a simple three-vowel inventory *i-u-a* with no harmony (but an innovative length contrast in the initial syllables).

3 Itelmen: quantifying the decay of a rule

In broad strokes, Itelmen is like Chukchi in the phonological aspects that are relevant to the current discussion. The basic vowel inventory is the same as that in Chukchi, as given in (13) (see also Volodin, 1976, 43).⁵

⁵ Volodin does not recognize schwa as a phoneme, suggesting that many instances of schwa are epenthetic (on which see also Bobaljik, 1998), while the remainder are highly reduced instances of other vowel phonemes, particularly in closed syllables with consonant clusters. Volodin’s orthography contrasts reduced and full vowels, but he explicitly treats only the full vowels as pho-

(13) Itelmen vowel inventory

recessive	i	u	e ₁
dominant	e ₂	o	a
transparent	ə		

Likewise, in terms of the alternations attested, Itelmen vowel harmony conforms to the Chukotko-Kamchatkan type: recessive vowels become their dominant counterparts in words with dominant vowels.⁶ The question that will occupy the remainder of this paper is a difference not in the phonological aspects of the vowel harmony process in Itelmen, but rather its pervasiveness in Itelmen grammar, and the striking change in the robustness with which alternations and exceptions are attested in materials over the course of three generations from 1910–1994.

The hypothesis I advance here is that Itelmen vowel harmony was (essentially) a productive process at the start of the 20th century and was effectively lost by the century's end. Qualitatively, this is relatively uncontroversial. Among the limited group of scholars who have considered the matter, there is a general consensus that Itelmen had a productive set of vowel harmony alternations, noted by Bogoras (1922) and visible in the texts recorded in 1910–1911, and that the contemporary language has a few alternating forms, but far less than it had a century ago. Writing about material he collected in the 1960s and 1970s, the preeminent Itelmen scholar A.P. Volodin already noted that the vowel harmony system (then still productive in nominal morphology) was losing ground in the verb:

Harmony is most inconsistently maintained in the finite verb. If the cases with the affixes *-a(t)* (which never controls [harmony]) and *-(xk)minj* (which never undergoes [harmony]) are put aside, it should be established that even the affixes represented by harmonic variants *-kičen ~ -kečan*, *-kinen ~ -kenan* et al. obey the demands of harmony in comparatively rare cases. Most often, the alternation does not take place: *tmaʔkičen* 'I played' (should have been: *tmaʔkečan*), *tk'otkičen* 'I came' (should have been: *tk'otkečan*), etc. The examples given above of harmonically regular verb forms look rather like exceptions. (Volodin, 1976, 46)

nemes. Bogoras (1922) gives an inventory with 22 vowels, although his contemporary Jochelson recognizes only 5; on which see below.

6 Although there are also some *i-a* alternations (Volodin, 1976, 43), as noted in n. 12 below.

Nevertheless, despite the substantial agreement on the broad pattern, the more interesting question I wish to pursue here is whether a quantitative evaluation can shed any light on the change, in particular, whether the type of approach to productivity pursued by Yang would show a critical change over this period. This section presents a preliminary investigation along these lines.

Note that the material from both periods (Jochelson's texts from 1910–1911, and mine from 1993–1996) have a mix of harmony-consistent alternating forms, and apparently disharmonic failures to alternate.

3.1 Contemporary Itelmen: 1993–1996

In the contemporary material, examples such as (14) and (15) show expected alternations. In (14), independently established recessive vowels (*i,u,e*) in roots change to their dominant counterparts (*e,o,a*) in the presence of a dominant suffix.

(14)	ROOT	HARMONY FORM	GLOSS	SOURCE
	ki(j)	ke -xʔal	river-ABLATIVE	A13
	isx	esx -anke	father-DATIVE	MimKp:2
	kist	kest -ank	house-DATIVE	Tilval:3
	kuke-	(x)an- koka -zo-nen	3.IRR-COOK-ITER-3>3SG	SP 47

In (15), we see the same alternation in an affix, triggered by a dominant root⁷:

(15)	AFFIX	ALTERNATING FORMS	GLOSS	SOURCE
	-enk	isx- enk	father-LOCATIVE	Tilval:2
		laχsχ- ank	mother-LOCATIVE	Tilval:2

Yet alongside these alternations, most affixes with recessive vowels fail to alternate:

⁷ The locative suffix *-enk~-ank* is recessive across Itelmen dialects. The related dative suffix is dominant wherever harmony-like alternations are attested. In the Khairjuzovo-Kovran dialect cluster, the dative is *-(an)ke*, as in *esx-anke* in (14). In the Sedanka-Tigil dialect group, the dative surfaces as *-(an)k* (e.g., *kest-ank*) thus creating a minimal pair with the locative in terms of their dominant/recessive behaviour. In Sedanka-Tigil, the dative also surfaces as diacritically dominant *-ŋ* (a borrowing from Koryak) or as *-ankəŋ*, combining the Itelmen and Koryak endings.

(16)	AFFIX	W/ DOMINANT ROOT	GLOSS	SOURCE
	-qzu	k-čāča- qzu -knen	PRT-cry-ASP-PRT	AS:1
	-βum	q-oms- qzu -βum-sx	2.IRR-leave-ASP-1.OBJ-2PL	AS:1
	-in	k'ot- in	come-3SG	S3:3
	-kičēn	n-alyt- kičē?n	1PL-spend.day-1PL	RasDan:50
	-kiłχ	elβant-zo- kiłχ	fish-ITER-NML	SP 22

Likewise, most affixes with dominant vowels fail to trigger harmony:

(17)	AFFIX	W/ RECESSIVE ROOT	GLOSS	SOURCE
	-kaq	siŋ -kaq	fly-NEG.PRT	AS:1
	-ał	qetit -ał-sx	freeze-FUT-2PL	AS:1
	-čax	jimsx -čax	woman-DIM	Tilval:1
	-laχ	ul'u -l'iaχ	little-ADJ	Tilval:1

Some roots fail to harmonize, even with affixes which do trigger alternations on other roots:

(18)	esχlin	esχlin -x?al	place.name-ABLATIVE	Tn:40
	kist	% kist -anke	house-DATIVE	(variation)

And there are numerous internally disharmonic morphemes:

(19)	zlatumx	sibling
	muza, tuza	1PL, 2PL PRON
	sinaŋewt, qusłnaqu	names (mythical figures) < Kor.
	niqa	quick(ly) < Kor. ?
	oxotil-	hunt < Russian

3.2 Itelmen in 1910

At first blush, a similar ambivalence characterizes Jochelson's material, and to some extent that collected by Jochelson's contemporary, Waldemar Bogoras.⁸

⁸ Bogoras (1922) specifically reports that the C-K vowel harmony system affects "almost all the vowels" (678) in Itelmen, yet his own examples include forms that appear to be disharmonic (even on the same page where he asserts vowel harmony applies), such as *k'ölkmin* 'he has come' (678), *tsünülotjk* 'I live in the woods' (679), etc. (*i,ü* recessive, *o,ö* dominant). In contrast to Jochelson's 5-vowel orthography, Bogoras gave 22 vowels for Itelmen.

The examples in this section are presented in Jochelson's orthography, which raises various issues to which we return below.

Examples such as (20) and (21) show expected alternations. In (20), independently established recessive vowels (*i,u*) in roots change to their dominant counterparts (*e,o*) in the presence of a dominant suffix.

- (20) isx 'father' **isx**-enk (LOC) **esx**-anke (DAT) K2.1
 kuke-'cook' **kuke**-ki (INFIN) **koke**-zo-xc (ITER-IMP)⁹ K2.27, 38

In (21), we see the same alternations in affixes, triggered by a dominant root:

- (21) -enk LOC isx-**enk** stó-al-**ank** K2.1, 5
 xonograf-**ank** K2.3
 -lax ADJ¹⁰ cíneŋ-**lex** caca-**lax** K2.10, 35
 íw-**lex** ás-**lax** K2.11, 4
 -(g)in 3SUBJ íł-**gin** łale-z-**en** K2.2, 5
 -kicen 1SUBJ t-pilgetí-z-**kicen** t-són-**kecan** K2.1, 2
 n-łxi-**kicen** n-ánta-**kecan** K2.3

While it appears that far more affixes undergo harmony in 1910 than in 1993, even in the older material, some affixes with recessive vowels apparently fail to undergo harmony:

- | (22) | AFFIX | W/ DOMINANT
ROOT | GLOSS | SOURCE |
|------|-------|--------------------------|----------------------------------|--------|
| | -ŋjin | hán-txal- ŋjin | 3.IRR-eat-3PL(>3) | K2.1 |
| | -in | k-tifsa-xk- in | CND-raise-II->2SG | K2.1 |
| | -min- | txál-a-s- min -sx | eat-DESID-PRES-1OBJ-2PL.
SUBJ | K2.4 |

And some affixes with strong vowels fail to trigger harmony:

⁹ Note that *e* seems not to alternate here; compare to the corresponding contemporary example above.

¹⁰ Volodin doubts that this affix alternated in Jochelson's time, despite these forms. See Volodin (1976, 76 n.25).

(23)	AFFIX	W/ RECESSIVE ROOT	GLOSS	SOURCE
	-ał	čki-ał-ki	find-FUT-INFIN	K2.39
		cf. čke-kaz	find-INFIN	K2.22 etc. (-kaz dominant)
		nú-ał-keq	eat-FUT-NEG	K2.23
		il-ał-c	go-FUT-2SG	K2.5
	-maŋ	q- téfsi-xk-maŋ	IMP-raise-II->1SG	K2.1

Some roots fail to harmonize, even with affixes which do trigger alternations on other roots (cf. (20))¹¹:

(24) *lexsx* **lexsx**-anke mother-DATIVE K2.1, 9

And there are numerous internally disharmonic morphemes:

(25)	<i>silatumx</i>	older sister
	<i>muza, tuza</i>	1PL, 2PL PRON
	<i>sinawet, kuskłiaqu</i>	names (mythical figures)
	<i>qula</i>	other
	<i>mozit</i>	is.able < Russian

3.3 The Jochelson-Danilov orthography

The examples above were presented in the previous section in the transcription used in Jochelson's material (Worth, 1961, 1969), about which a number of remarks are in order before we proceed to a count.

The Jochelson collection comprises 41 texts of varying length (conventionally numbered K2.1, K2.2, etc.), in total 277 pages (including translations) in the published version (Worth, 1969), collected in Kamchatka in 1910–1911. The majority are from the southern (Khairjuzovo) dialect, with a few from the northern (Sedanka) variety. There are hints that the texts were transcribed by a native Itelmen speaker working as a guide and assistant to Jochelson, probably one A.M. Danilov (see Bobaljik & Koester, 1999). Worth (1969) compiled a dictionary from the texts, which serves also as a partial concordance, listing all distinct word-forms. The dictionary, including headwords and examples, has 4,285 wordforms

¹¹ The vowel in *lexsx* is not dominant, cf. vocative *lexsx-e* rather than *lexsx-a* and discussion at (38) below.

(tokens). Of these, 861 (20%) are on the face of it violations of vowel harmony, containing at least one unambiguously dominant vowel $\{o,a\}$ and at least one unambiguously recessive vowel $\{i,u\}$. (Since e may be either dominant or recessive, it was excluded from this count.)

Closer scrutiny of the texts shows that it would be misleading to conclude that this represents the actual number of disharmonic forms, for a few reasons.

Notably, more than 40% of these exceptions (371 of the 861 exceptions) involve an a between a uvular and a sonorant (including $/z/$), in a word with otherwise recessive vowels, as in the following:

- (26) a. ksunɫqazúknen
 b. qaziɫqazuknen
 c. kúneɲtqazuknen
 d. kkelqazúknen

The “a” here is undoubtedly excrescent: a brief, but audible release of the uvular stop before the following voiced segment. Contemporary transcriptions, such as Volodin (1976) and my own notes, do not indicate this as a vowel, and it is not consciously perceived by speakers as such (although it is still audible as a transitory element). Contemporary transcriptions corresponding to (26a-b) are given in (27).

- (27) a. k-sunɫ-qazú-knen Jochelson
 k-sunɫ-qzu-knen Contemporary
 PRT-live-ASP-PRT
 ‘He lived.’
 b. qaziɫ-qazu-knen Jochelson
 [k]-qziɫ-qzu-knen Contemporary
 PRT-get.ready-ASP-PRT
 ‘He got ready.’

In this context, it is worth noting that Jochelson (or Danilov) frequently indicates an accent on the vowel immediately following this excrescent vowel, as in three of the four examples above. There is no interpretation of stress in Itelmen for which this would make sense, but we can presumably understand it as indicating the relative perceptual contrast between an excrescent and adjacent full vowel.

Many of the excrescent vowels, as in the examples in (26), appear in the aspectual morpheme $/-qzu-/$. In the 1910 texts, this morpheme behaves regularly as concerns vowel harmony, if one ignores the excrescent vowel, alternating between $-qazu-$ and $-qazo-$, as in (28). These examples all show vowel harmony

behaving as expected not just with the aspectual morpheme, but with subsequent inflectional suffixes as well, once the excrescent *a* is factored out.¹²

(28)	RECESSIVE ROOT		DOMINANT ROOT
	k-sunł- qazú-knen		k-wetat- qazó-knan
	k-txzi- qazú-knen		k-xaimanto- qazó-knan
	min-sxezí- qazu-sx		k-swatal- qazó-knan
	k-tmpl- qazú-in		k-tpal-zo- qazo-án

Two other characteristics of the transcription scheme result in apparent disharmony, where there probably was none. Many examples of orthographic “i” in 1910, especially those adjacent to another vowel, are undoubtedly glides /j/, and thus not subject to the harmony process:

(29)	Jochelson:	a(y)iwa	káitatān	csalai	brawoi
	contemporary:	aʔjuβʔaj	kʼ-ajtat-an	tsal-aj	braβ-oj
		brains	herded	fox-AUG	good < Russian

In addition, Jochelson-Danilov uses a five-vowel transcription system. Many vowels written with full vowel characters in 1910 correspond to schwa in the contemporary language, which has (at least) five full vowels plus schwa.¹³ Some examples are given in (30):

(30)	Jochelson:	ína	kíma	kantxigaan
	Contemporary (S):	ənna	kəmma	k-əntxa-(ʔ)an ¹⁴
		3SG.PRON	1SG.PRON	PRT-forget-TR.PRT

It is implausible that the discrepancy represents a change – reduction from full vowels to schwa – over the last century. Schwa is a prominent part of the vowel inventory in all of the Chukotko-Kamchatkan languages, and for some of the items

¹² The “i” in *xaimanto-* is a glide, not subject to harmony; see below. In general, harmony alternations in Chukotko-Kamchatkan pair *i-e* or *e-a*. There is a small number of alternations in Itelmen which instead pair *i-a*. The alternation *-in--an* for the transitive participle (in the last line of (28)) is one such example, and is preserved to some extent in the modern language. The cognate ending shows an *e-a* alternation in the Chukotkan languages, perhaps suggesting that *i-a* arises as a means to avoid the neutralization that would result from the *e-a* merger in some Koryak dialects.

¹³ The distribution of schwa is largely, but not entirely, predictable. See Bobaljik (1998)

¹⁴ Jochelson’s “g” = [ɣ] was preserved in inflectional morphology in the speech of the speakers who consulted for Volodin (1976), although few traces remained in the 1990s.

in (30), the schwa is consistently present in the cognates in the other Chukotkan languages, e.g., Chukchi 3SG.EMPH/REFL *ənan*, 1SG *ɣəm*, strongly suggesting that these words did not have full vowels in the previous generation in Itelmen, and that this is indeed an artifact of the notation.¹⁵

In addition to the systematic factors just indicated, there is some measure of internal inconsistency in the Jochelson-Danilov transcriptions, as there is in all subsequent corpora of any substantial size. Presumably, we may recognize some measure of noise introduced by transcriber error and other factors.¹⁶

Factoring out excremental *a*, glides, and presumed schwas leaves us nevertheless with a residue of items which appear to be disharmonic, many consistently so across multiple occurrences in the texts, making transcriber error unlikely. Some examples are given here. Those in (31) are probably loan words from Russian, or from Koryak *a*-dialects (recall from above that the *a*-dialects of Koryak and Alutor have seen recessive *e* merge with *a*; this recessive *a* does not undergo harmony and triggers only *e* → *a*, but leaves *i, u* unaffected):

- (31) *docista* < Russian: *dočista* ‘clean / everything’
mozit < Russian *možet* ‘is.able’
ilyá < Russian *ilja* (name)
sinaŋewt < Koryak *jiniŋawɣut* (name, mythological figure)
 (s : j is regular)
kuskłiaqu < Koryak *qujqinjaqu* (name, mythological figure)

But there are also disharmonic words that are less obviously¹⁷ loanwords, such as those in (32):

¹⁵ More so than the considerations above, identifying a particular vowel in the Jochelson-Danilov transcriptions as schwa requires some measure of guesswork, both since there is some variation in the contemporary language, and in some cases, since the corresponding contemporary form is not readily identifiable. In addition to my own field notes, I have made use of Volodin & Khaloimova (1989) and the modern edition of Jochelson’s tales, as edited by K. N. Khaloimova (Khaloimova et al., 2014).

¹⁶ From personal experience, I can attest to blurry boundaries in the vowel space, for example between *i* and *e*, or *e* and *a*, leading to uncertainty especially among non-native transcribers, but even among native speakers aiding in editing and transcribing.

¹⁷ Where there are apparent cognates, as in the first person plural pronoun, it is of course difficult to determine whether these are loanwords, or true cognates representing a common ancestor. Since I will not be excluding loanwords in the counts, settling this challenging point is not necessary.

- | | | | |
|------|-------|---------------|--|
| (32) | muza | 1PL.PRON | cf. Koryak <i>muri</i> |
| | qulán | ‘other’ | cf. Koryak <i>qul(i)</i> , <i>qullu</i> , also Kerek <i>qula</i> ¹⁸ |
| | awi | ‘crab’ | cf. Koryak <i>avi</i> |
| | akiká | INTERJ (Hot!) | |

Finally, of particular note is the FUTURE/DESIDERATIVE suffix *-at*, which is inert to harmony both in the 1910 material and in contemporary forms, and which curiously has no known cognate in Chukotkan. Examples of this affix failing to trigger harmony were given in (23); an additional minimal pair with a recessive and dominant root (with harmony applying across the future affix) are given in (33):

- | | | |
|------|--------------------------|---------------------------|
| (33) | RECESSIVE ROOT | DOMINANT ROOT |
| | t-lxiln-ál- kicen | ta-wetat-al- kecan |
| | 1SG-stop-FUT-1SG (K2.1) | 1SG-work-FUT-1SG (K2.30) |

In sum, while there are surface disharmonic forms in the Jochelson texts, they are not nearly as pervasive as a superficial count of vowels as written might suggest. Taking account of the various important quirks of the Jochelson-Danilov transcription, we may proceed to a quantitative evaluation of samples of Itelmen from the beginning and end of the 20th century.

4 Counting harmony

At both the beginning and the end of the 20th century, the Itelmen child was faced with ambivalent evidence for the harmony rule in (34), where [D] is [-ATR] or whatever feature it is that relates pairs of dominant and recessive vowels, and the rule is read without regard to linear order: a recessive vowel is changed to its dominant counterpart in a word with a dominant vowel.

- (34) $V^{[-D]} \rightarrow V^{[+D]} / [\{ \dots _ _, V^{[+D]} \dots \}]_{\omega}$

At both time periods, some vowels alternated, but some did not, and in both periods, there were superficial exceptions to harmony – words that on the surface contain a mix of dominant and recessive vowels. By all accounts, (34)

¹⁸ From Fortescue (2005).

(or something similar) was acquired as a productive rule by speakers who lived a century ago, but not acquired by the current generation of speakers. Why?

4.1 Tolerating exceptions: Yang's threshold

Yang (2016) proposes a means to determine whether a rule is learnable in the face of exceptions. Intuitively, what Yang proposes is a measure of the threshold of permissible exceptions – the cut-off point for deciding whether the proper analysis is to posit a rule, with some exceptions, or whether it is more appropriate to simply list the alternating forms lexically. Specifically, (Yang, 2016, 64) proposes the Tolerance Principle in (35), which defines the cut-off point for the learnability of a rule¹⁹:

(35) *Tolerance Principle*

Let R be a rule applicable to N items, of which e are exceptions. R is productive iff:

$$e \leq \theta_N, \text{ where } \theta_N := \frac{N}{\ln(N)}$$

In this formula, N is the number of opportunities for the rule to apply (instances where the structural description is met), and e is the number of exceptions. The formula counts types, not tokens and thus a few high-frequency exceptions will not undermine an otherwise productive rule. In Yang's analysis, the regular past tense inflection of English (add *-ed*) is productive, despite the existence of exceptions, but to a first approximation, irregular forms must be learned on an item-by-item basis, despite subregularities.

Table 1 reports the results for a count from this perspective for the first text from Jochelson's collection, and the first text from my collection. The first line of the table is the total number of distinct words in each text. This line counts types, not tokens, and is not lemmatized (since a given root may combine with dominant or recessive inflectional affixes). N counts the number of words (types) whose underlying representation contains at least one unambiguously

¹⁹ The following discussion is not meant to necessarily endorse the idea of a sharp dividing line between productive and non-productive rules, but to ask whether it is possible in principle to quantitatively characterize such a divide in a way that makes sense of the Itelmen change. I am sympathetic, in principle, to the idea that a more articulated model might assume that learners consider multiple potential rules, with weighted probabilities (as in Albright & Hayes, 2003), and that there may be a gray area near the threshold.

Table 1: Yang's measure of productivity for vowel harmony in two Itelmen texts.

	K2.1 (1910)	Angaqe (1994)
words	242	229
N	49	42
e	11	37
Yang's Threshold θ_N	13	11
Productive ($e \leq \theta_N$)?	Y	N

dominant element and at least one unambiguously recessive element. e counts the number of these which fail to resolve the harmony conflict and remain disharmonic on the surface.²⁰ Unsegmentable interjections (*akika(x)*! 'Ouch! [for something hot]') and (probably borrowed) proper names (*Sinaŋewt*, *Ilyá*) are excluded, on the grounds that it is not uncommon for these to fall outside the regular phonology.²¹

20 A reviewer raises the question of whether this is the correct choice for N , asking whether N should include all polysyllabic words, including those which are underlyingly harmonic. Clearly, this would increase N (and thereby also θ_N) in Table 1, without increasing e , and would thus affect the key calculation. If I have understood correctly, the reviewer's suggestion asks a different question from Yang's. The reviewer's suggested approach to N asks about surface distribution: to what extent is the set of surface forms consistent with the constraint in (i) (a ban on words containing a mix of dominant and recessive vowels, i.e., the constraint corresponding to the rule in (34)):

- (i) *{ ... $V^{[-D]}$, $V^{[+D]}$... }] ω

An approach along these lines is taken by Harrison et al., 2002; Dras & Harrison, 2003—see fn. 30 for some discussion. My understanding of Yang's calculation, and thus the approach to N taken in the text, is narrower. It asks how often the rule in (34) applies in contexts in which it could apply. The assumption is that evidence for a rule comes only from alternations in which the rule actually applies. Just as words with only a single (full) vowel provide no evidence for whether there is a harmony rule (although they trivially satisfy (i)), words with underlyingly harmonic vowel combinations do not distinguish between a grammar that has a harmony rule (or constraint) and one which lacks such a rule, and are thus not counted as instances of N . In a fuller treatment of this material I hope to explore in more detail the consequences of differing assumptions for the model, but must leave this for future work.

21 The corresponding figures with names and interjections included are: K2.1: $N = 54$, $e = 16$, $\theta_N = 14$ and Angaqe: $N = 45$, $e = 40$, $\theta_N = 12$. While there is still a substantial difference between the two time periods, on this way of counting, the exceptions to the harmony system would fall slightly above the threshold even in 1910. I provisionally take this to be an artifact of counting words, rather than morphemes: the single disharmonic name *Sinaŋewt* shows up in three

By Yang's criterion, vowel harmony was learnable as a productive rule in 1910, and unlearnable three generations later (assuming that these results scale up to the larger corpus). The two texts are roughly the same size, and contain comparable values for N (and thus for Yang's Threshold θ_N), but differ drastically in their values for e —the number of surface exceptions to harmony. Even allowing for some degree of transcriber error in identifying vowels in the *Angaqe* text, the number of exceptions there exceeds the threshold by a factor of three.

While this result provides quantitative support for the initial hypothesis that vowel harmony was productive in 1910 and has been lost recently, there are various qualifications worth making in understanding the numbers.²² In the next paragraphs, I discuss some of the practical choices that went into the count, acknowledging that at each point, different choices could have been made, perhaps leading to different results.

The first point to note is that I have, for the purposes of this pilot study, followed Yang in counting word-types in the texts. For Yang's examples, such as the English past tense or the German plural suffixes, the combinations of interest are bi-morphemic words. For each English verb root (or stem) there is in the general case one word which constitutes the past tense of that verb. Thus words (types) are a convenient proxy for morpheme combinations that do or do not trigger a specific rule. But Itelmen words, in particular verbs, are significantly more complex, often multi-morphemic. The form in (36) (from the representative 1994 text in the table) has three vowels in five morphemes, and on the surface is an exception to vowel harmony. By counting words, this word adds one each to the value of N (the harmony rule could have applied) and e (an exception).

- (36) q'-oms-qzu-βum-sx
 2.IRR-leave-ASP-1SG.OBJ-2PL.SUBJ
 'You (should) leave me.' (AS)

different word-types in this short text; this name alone effectively makes the difference between productive and non-productive results in this small sample.

²² There is also, as a reviewer notes, an important question of what productivity means in general and how we establish it. Since there are morphologically-specified exceptions to harmony in all attested Chukotko-Kamchatkan languages, one might wonder if it is ever accurate to call the process productive. Table 1 documents a substantial change in the degree of harmony-like alternations in Itelmen words, which (as we will see below) correlates with the rise of Russian loans over the same time period, many of which are disharmonic. Yang's theory provides an explanatory mechanism for understanding this correlation in terms of the loss of productivity — the eventual unlearnability of a previously learnable productive rule. It is, however, true that no systematic *wug*-test was conducted at any point over the history of Itelmen, and is infeasible now.

On the other hand, there are two morphemes with recessive vowels in this word, neither of which changes. One could instead have counted morphemes or vowels, but this greatly increases the complexity of the exercise. If we count morphemes (or vowels), should (36) count as two exceptions, since two recessive vowels fail to undergo the rule, or one, since the single dominant vowel fails to act as a trigger? The task is even more complex with pairs like the following, from the same text:

- (37) a. *seŋ-zo-z-in* b. *k-siŋ-qzu-kne'n*
 fly-NDIR-PRES-3SG PRT-fly-ASP-PRT.PL

The root /*siŋ*/ ‘fly’ (seen as such in (37b)) surfaces as *seŋ* in (37a) under the influence of dominant *-zo* (a derivational morpheme that derives the non-directed meaning ‘fly around’ from ‘fly’). Thus harmony has applied in this form, but at the same time, it remains a surface violation, since the inflectional suffix *-in* retains its recessive vowel. By counting words, (37a) counts as an exception, despite the harmony alternation in the root. Note that this increases the proportion of $e : N$, making it harder for a productive rule to be detected as such.

Another way in which counting words under-counts N is that a word that shows multiple instances of the harmony rule applying (such as *k-caqał-qzo-knan* </*k-caqał-qzu-knen*/) will contribute only one N to the sum. By under-counting harmony rule application in this way, we are, if anything, setting Yang’s Threshold too low, and thus we can be that much more confident in a positive result for productive harmony.

On the other hand, counting word-types, rather than morpheme types, runs the risk of allowing high frequency morphemes to have an inordinately large effect of the outcome. But this cuts both ways. The most frequent morphemes²³ in Jochelson’s K2.1 text include the aspectual morpheme *-qzu*, which alternates with *-qzo*, and the future morpheme *-ał* which, as noted above, fails to trigger harmony as it should.

Establishing the value for N requires a determination of the underlying representation of each form in the texts. Each word in each text was segmented by hand into constituent morphemes, and its UR was determined as accurately as possible by comparing to other occurrences of the same morpheme. This is not always trivial, and is a source of possible inaccuracies, in particular in cases where a given morpheme is infrequent. The treatment of surface e is particularly tricky. Two words from K2.1 are given in (38); each is bi-morphemic and both contain 2 instances of the vowel e (and no other vowels):

²³ I.e., those that occur in the most distinct word types.

(38) Surface:	a. <i>lexsx-e</i>	b. <i>sen-ke</i>
UR:	/lexsx-e/	/sin-(an)ke/
gloss:	mother-voc	woods-DAT/ALL
	‘O mother!’	‘Into the woods.’
source:	K2.1	K2.1

Surface *e* may correspond to any of: (i) dominant /e/, (ii) underlying /i/ after the application of harmony, (iii) recessive /e/ in a non-harmony context, or (iv) an exceptional, non-alternating, but also non-dominant /e/. As it happens, (38) shows all four. Starting with (38b), the (bound) root for ‘forest, woods’ is /sin/, seen in its recessive form in *sin-k* ‘forest-LOC’; in (38b), the vowel has changed under the influence of the dative/allative suffix *-(an)ke*. This suffix has a long and short form, and the short form *-ke* is a reliable harmony trigger, as in *ηon-ke* ‘here-ALL’ < *ηun* ‘here’. In (38a), the vocative suffix /-e/ is recessive and alternates with [-a] after a dominant stem, as in *la:ηé-sg-a* ‘girl-PL-VOC’. This could suggest that *e* in the root *lexsx* ‘mother’ is also recessive, however, this root fails to alternate in the Jochelson texts, and stays as *e* before dominant affixes, as in *lexsx-anke* ‘mother-DAT.’²⁴

As these examples indicate, recovering the underlying form is not always a simple matter. Surface *e* is compatible both with dominant and with recessive contexts, and some amount of analysis is required to identify both the action of harmony (as in *sen-ke*) and apparent exceptions (such as inert, non-alternating, non-triggering *lexsx*). For practical reasons then, underlying representations of morphemes were counted as unambiguously recessive if either (i) they contain {i,u} or, if their vowel is *e* and they demonstrably alternate with *a* in the corpus (such as vocative *-e~-a*, participial *-knen~-knan*). Likewise, unambiguously dominant forms are those with {a,o}, or those which clearly behave as such (such as the short dative allomorph *-ke*, as in (38)).

Although setting aside non-alternating *e* in this matter may reduce the accuracy of the counts, another interpretation of this is that the phonology of vowel harmony in 1910 was already somewhat different from the general Chukotkan pattern. Rather than a division of all vowels into recessive and dominant pairs, perhaps already in 1910, it was better to think of a recessive series *i,u,e*, a dominant series *a,o*, and a neutral series *e₂,ə*. On this view, there is no ‘dominant *e*’; the apparent dominant behaviour of the short dative/allative *-ke* is then attributed to a morpheme-level diacritic, rather than the vowel itself. Note in this context that the dative is a dominant affix also in Chukchi and Koryak, where it has no full

²⁴ In contemporary materials, the attested form is *laxsx* or *laxsx*.

vowel and must be marked with a diacritic: Chukchi *nenenə* ‘child’, *nanan-ɣtə* ‘child-DAT’ (Kurebito, 2012, 182), Koryak: *milute-k* ‘hare-LOC’, *melota-ŋ* ‘hare-DAT’ (Abramovitz, 2015, 3). From this perspective, the behaviour of *lexsx* is not in any way exceptional, and it is right to therefore exclude it from the count, as I have done.

Here, one could consider again the view of Chukotkan vowel harmony in Calabrese (1988) in which the feature [+D] in the vowel harmony rule (34) is always a morpheme-level diacritic. From this perspective, exceptions above are not exceptions to the vowel harmony rule as such, but rather to the generalizations that regulate the assignment of the diacritic [+D]. In addition to lexically-specific marking of [+D] for certain morphemes (required in any event for dominant morphemes with no full vowels), there would be a rule that assigns the diacritic [+D] to any morpheme that contains /a/ or /o/:

$$(39) \mu \rightarrow \mu^{[+D]}/[\dots \{a, o\} \dots]_{\mu}$$

The (older) Chukchi analog of this rule would have included dominant *e* (assuming it was at some point phonetically distinguishable from recessive *ɛ*) in the context for assigning the [+D] diacritic.

Despite these qualifications, it is not seriously in doubt that there has been a qualitative change in whatever vowel-alternation processes Itelmen has, and that this change has been significant over the course of the 20th century. Even with many sources of uncertainty, the major finding reported here is that Yang’s Tolerance Principle appears to quantitatively bear out the hypothesis that there was a productive process of vowel harmony operative in 1910, and that there no longer is.

4.2 Russian loans

The previous sections tentatively establish that there has been a substantive change in the vowel harmony pattern in Itelmen over the course of the 20th century, but the numbers alone do not provide any indication of the causal mechanism involved. The loss of vowel harmony is presumably an effect of language shift, a product of the declining spheres of influence of Itelmen. Russian subjugation of Kamchatka began in the 18th century, and continued throughout both the tsarist and Soviet periods. As the numbers in (40) show, not only was the ethnically Itelmen population decimated, but the number of Itelmen people who retained Itelmen as their native language contracted substantially:

(40) Demographics

	ca. 1700	1926	1994	2001	Sources
ethnic speakers	10–25,000	3,414	1,141		Stebnickij (1934); Volodin (1976)
	all	803	<80	<40	Koester & Bobaljik (1994)

By the time of Jochelson's expedition in 1910–1911, Itelmen was spoken only in 8 villages on the remote Okhotsk Coast. The Soviet period saw further drastic decline: schools were established, and children were prohibited from speaking Itelmen, in some case forcibly removed to boarding schools. Itelmen villages were razed and Itelmen speakers forcibly resettled. By 2017, a veritable handful of fluent, native speakers remain.²⁵

Language shift alone does not explain the loss of harmony. I offer here the hypothesis that a more specific mechanism was the large influx of disharmonic Russian loanwords into the quotidian vocabulary. Although Table 1 shows that harmony was productive in 1910, it was only marginally so – the number of exceptions was only just below Yang's Threshold. Consider in this light the counts in (41). This table counts the number of Russian loan words in the entire Jochelson corpus as compared to two selected texts from 1994:

(41) Loan rates (Russian words / Total words) [lexeme types]:

CORPUS	# lexemes	russian	loan rate	notes
1910 – Jochelson	1546	130	8.4%	entire corpus
1994 – Tilval	243	48	20%	youngest fluent generation
1994 – KL	279	50	18%	youngest fluent generation

The number of Russian loans more than doubled. Although I do not have a count at this time of what percentage of Russian loans are disharmonic, it can be observed that it would take only a few to push the number of exceptions over Yang's Threshold.

Russian loans represented 8% of the vocabulary in 1910 and as much as 18–20% by 1994. For a text comparable to K2.1, we would expect to find perhaps

²⁵ There is room for differences in the criteria, but there are currently fewer than 5 speakers for whom Itelmen was their first language, and who continue to speak it fluidly.

15–20 more Russian loans in 1994 than in 1910.²⁶ Even if a mere 3 wordforms among these loans were disharmonic (and did not ‘displace’ equally disharmonic wordforms), the rate of exceptions would surpass Yang’s Threshold and make the critical change from a learnable to an unlearnable rule. While a missing variable in this back-of-the-envelope calculation is the actual percentage of the Russian loanword vocabulary that is disharmonic in Itelmen, the result just mentioned is that it would be enough if only roughly 1 in 5 to 1 in 7 Russian loans are like those in (42) (all attested in the 1993–1994 texts) having at least one *a,o* and at least one *i,u*:

- (42) a. bumag(aʔ) ‘paper’
 b. babu-čχ ‘grandmother/old.woman-DIM’
 c. izmennoj ‘betrayed’²⁷
 d. oxotił- ‘hunt’ (k-oxotił-qzu-knen ‘PRT-hunt-ASP-PRT’)
 e. natjanut ‘draw’ (a bowstring)

By no means does this conclusively establish that Russian loans were the proverbial camel’s-back-breaking straw, but I suggest that the figures above establish a clear quantitative case for the plausibility of the scenario entertained here. From the numbers, we have seen that the number of exceptions to vowel harmony in Itelmen in 1910 was perilously close to the critical threshold beyond which the rule could no longer be learned; a mere handful of disharmonic loans would be enough to push the harmony process over that cut-off. We know that many Russian loans are disharmonic, and we know that the rate of Russian loans occurring in the texts increased substantially over the relevant period.

If this hypothesis is on the right track, it also has implications (as a reviewer observes) for how lexical stratification and productivity interact. In theory, one could imagine that the Russian loans constitute a distinct lexical stratum or co-phonology, (cf. Ito & Mester, 1999) and that the harmony rule is limited to the native stratum. Yang recognizes that rules that may seem non-productive over the whole language may emerge as productive in discrete sub-domains. If learners were to have entertained the hypothesis that the harmony rule is limited to the

²⁶ The text K2.1 has 242 word (types), but this count includes different inflected forms of the same lexeme; where the figures in (ii) count lexemes. Using the same criteria, K2.1 has approximately 175 lexemes. If 8% were Russian loans in 1910, these would number 14 lexemes; where we would expect 31–35 loanword lexemes under the 1994 loanword rate.

²⁷ Borrowed Russian adjectives in Itelmen tend to be borrowed with the fixed adjectival suffix *-oj* regardless of the suffix formative in Russian.

native stratum, they may discount un- or partially-assimilated Russian loans as being irrelevant to the productivity of the rule. In Bobaljik (2006) I argued that the Itelmen lexicon is indeed stratified, and that reduplication is limited to the core stratum, which is distinguished by a number of other properties, while the non-native vocabulary spans a variety of strata. However, there is evidence that harmony extended to the stratum or strata containing Russian loans inasmuch as they do (or did) participate in harmony, at least as triggers – borrowed roots with dominant vowels do occur with the dominant versions of alternating affixes, as shown in (43)²⁸:

- (43) a. kápusta-ʔaʔn ‘cabbage-ATTRIB.PL’ < Russian *kapústa* ‘cabbage’
 stol-ank ‘table-LOC’ < Russian *stol* ‘table’
 b. k-swatał-**qazó-knan** ‘PRT-WOO-ASP-PRT < Russian *svatat*’ ‘woo’
 Nówoi-gód-ank ‘New-Year-LOC’ < Russian *Novyj God* ‘New Year’

Evidently, lexical stratification as such does not necessarily prevent loanwords from being taken into consideration in the computation of the productivity of a rule. A related conclusion is drawn in Drescher & Lahiri (2015) who offer an analysis of the influence of Romance loanwords on English stress, also couched in terms of Yang’s Threshold. Their analysis of English differs somewhat from the tentative hypothesis I have advanced here, for example, in that it is not the absolute number of Romance loans with a non-English stress-pattern that matters, but rather the evidence for stress-shifting latinized affixes that affects the learnability of the stress rules differentially over the history of English. But their account shares with the one offered here, and indeed with any account that pins phonological change on the influence of loanwords, the view that lexical stratification does not automatically exclude loanword strata from the computation of Yang’s Threshold.

5 Beyond Itelmen: disharmony in a broader context

Before closing, I offer a final speculation. Other languages have seen an influx of foreign, disharmonic loans, and yet retain vowel harmony systems. Turkish is a widely discussed case in point (Clements & Sezer 1982). Why should the outcome be different in different languages? After all, Yang’s Tolerance Principle regards

²⁸ The a. examples are from my fieldwork, and the b. examples from Jochelson’s first two texts.

the proportion of exceptions in (a corpus representative of) the Primary Linguistic Data available to the child settling on their grammar. I speculate here that one critical difference is in the nature of the harmony process itself, and thus the rule to which Yang's learnability calculation applies.

Turkish vowel harmony is illustrated in (44)–(45) (examples from Kabak, 2011, which offers an overview). The basic harmony system is that vowels agree in [backness] with the immediately preceding vowel (as in (44)), and high vowels agree moreover in roundness (as in (45)).

- (44) a. dal-**lar-un** branch-PL-GEN 'of branches'
 b. jer-**ler-in** place-PL-GEN 'of places'
- (45) a. boyn-**un** 'neck-2SG.POSS'
 b. gøuṣ-**yn** 'breast-2SG.POSS'
 c. aln-**un** 'forehead-2SG.POSS'
 d. vakt-**in** 'time-2SG.POSS'

There is, however, a notably large inventory of disharmonic roots, many of which are loanwords from Arabic and from European languages (Clements & Sezer, 1982; Kabak, 2011):

- (46) kitab 'book', siroz 'cirrhosis', garip 'strange', polis 'police',
 butik 'boutique', pilot 'pilot'

For the sake of argument, let's assume the number of disharmonic roots in Turkish is comparable to Itelmen. So why has this inventory not brought down the harmony system in Turkish?

As various authors have noted, the generalizations about Turkish are of two types. Affixes undergo alternations: essentially every suffix in Turkish has multiple surface allomorphs, and thus there is a motivation for a phonological harmony rule governing these alternations. But roots in Turkish do not undergo alternations. To the extent there are generalizations about the distribution of vowels in Turkish roots, these are Morpheme Structure Constraints (MSCs, see Kiparsky, 1973), static generalizations about lexical items which are not the result of a phonological rule. One way of approaching this (Clements & Sezer, 1982) is to posit that affix vowels are unspecified for backness and roundness, while roots, learned directly as such from the output, are fully specified. The harmony rule on this perspective is a feature-filling rule. Whatever technical implementation is chosen, assimilating 'disharmonic' loanwords into Turkish for the most part is then a matter of updating the lexicon. The feature-filling harmony rule is unaffected by roots that do not conform to the historically motivated MSCs. Indeed,

whether a root is disharmonic or not, the final vowel of the root typically spreads its backness (and, to high vowels, roundness) rightwards, as in (47), participating in the productive part of the harmony system:

- (47) a. kitab-**un** book-GEN
 b. siroz-**un** cirrhosis-GEN

In other words, because Turkish vowel harmony is directional, and thus root-controlled (since Turkish is a suffixing language), apparently disharmonic roots can be easily integrated into the phonological system of Turkish, without affecting the phonological rule that governs alternations (and thus without jeopardizing its learnability).²⁹

Chukotko-Kamchatkan is crucially different from Turkic in this regards. In Chukotko-Kamchatkan, harmony operates in both directions and roots, as well as affixes, undergo alternations. A borrowed ‘disharmonic’ root in Turkish is, as we have seen, not only not a violation of the productive (rule-governed) part of harmony, it participates fully in that system, in light of the underlying representation of its final vowel. By contrast, a disharmonic root in Chukotko-Kamchatkan truly is an exception – no matter how it interacts with affixes, it will either contain an apparently dominant vowel that fails to act as a harmony trigger, or it will contain a recessive vowel that fails to act as harmony undergoer. This line of reasoning offers a straightforward account of why the Itelmen harmony system would be more vulnerable to decay under influence of loanwords than a root-controlled system would be. What remains to be shown is that this hunch scales up. Various languages with root-controlled systems are known to have lost harmony (see for example, Harrison et al., 2002; Dras & Harrison, 2003 on some Turkic languages, also Estonian), and Chukchi, by all accounts, has retained the bi-directional dominant-recessive system, despite undergoing language shift towards Russian and thus incorporating Russian loans (see Dunn, 1999).

An approach such as Yang’s gives us a potential tool with which to investigate these differences systematically.³⁰ The conjecture just given is that the effects of

²⁹ There are also a few disharmonic suffixes in Turkish, but the key contrast with Itelmen is in the role of roots in the harmony system.

³⁰ Harrison et al. (2002); Dras & Harrison (2003) provide a different quantitative approach to the learnability of vowel harmony, with specific attention to variation across Turkic. In a nutshell, their approach counts the relative frequencies of each vowel in a corpus of data, and calculates from this the expected distribution of combinations of vowels in polysyllabic words. In the simple case, if there are two equal groups of vowels (dominant and recessive) that could be combined freely, then 50% of bisyllabic words should be ‘disharmonic’, containing one vowel

disharmonic roots will be more pernicious in bi-directional, root-affecting systems than in systems of root-controlled spreading. But the force of the exceptions is relative to the overall robustness of the evidence for the rule. Independent of the Russian loans, all of Chukotko-Kamchatkan vowel harmony has some unpredictable aspects that require a diacritic mark (such as the dominant morphemes that lack a full vowel). And within Chukotko-Kamchatkan, Itelmen, much more than Chukchi, evidences other surface exceptions to the harmony system: the ambivalent behaviour of *e*, reflexes of the *a-e* merger from neighbouring Koryak *a*-dialects, and indigenous, harmony-violating morphemes, such as the harmony-inert future *-at*. It strikes me then, as not out of the question that a careful, quantitative comparative analysis from the perspective taken here may turn out to be able to characterize the finer distinctions in the preservation or decay of vowel harmony under language contact, and to support or refute the conjecture offered here, that bi-directional systems in which roots alternate will be more vulnerable in language shift than root-controlled, directional harmony.

Acknowledgments: Above all I am grateful to the many speakers of Itelmen who have shared their language with me in Kovran, Tigil, and Sedanka (Kamchatka, Russia) since 1993. The text analyzed in detail in section 4.1 was recorded in Tigil in 1994, as told by the late Tatiana N. Bragina, and transcribed with the generous help of Ludmila E. Pravdoshina. This paper has had a long gestation. I am glad to be able to contribute it to this volume, in recognition of many stimulating and enjoyable discussions with Andrea Calabrese over the years we have been friends and colleagues, including conversations about some of the material here. In addition to Andrea's feedback, I have benefited from discussions at various points with Rafael Abramovitz, Jeff Bernath, Michael Krauss, Irina Monich, Andrew Nevins, †Aleksandr P. Volodin, Susi Wurmbrand, Charles Yang, and audiences at the LSA (1995), McGill University, the University of Connecticut, and the Workshop on the

from each group. If, however, recessive vowels were significantly more frequent to start with, then the basic statistical computation would yield a higher proportion of apparently 'harmonic' words, in which all vowels are recessive, just because most vowel tokens are recessive. Harrison et al. (2002) identify a harmony process as learnable if the distribution of vowels in multi-vocalic words is more harmonic than would be expected from combining the baseline frequencies of the individual vowels. Under their system, even contemporary Itelmen has vowel harmony, evidently a false result. Although I leave demonstration of this to a more fuller presentation at a later date, the source of the inadequacy of the Harrison et al. model lies in the effect of alternations. Any vowel alternation will skew the attested combinations of vowels away from baseline chance and the system has no way to distinguish between a productive phonological rule and morphologized, listed ablaut-like alternations.

Division of Labour between Morphology and Phonology at the Meertens Institute (2009). I also thank two anonymous referees for their valuable comments. For assistance with various text counts, I thank Ksenia Bogomolets and Irina Monich. Errors are, of course, my own responsibility. Funding for portions of the research reported here was provided by the National Council for Soviet and East European Research (PI: D. Koester), and NSF BCS-1065038 and BCS-1263535.

Abbreviations: The following abbreviations are used here:

ADJ	adjective	ALL	allative	ASP	aspect
ATTRIB	attributive	AUG	augmentative	CND	conditional
DAT	dative/allative	DESID	desiderative	DESIG	designative
DIM	diminutive	FUT	future	II	class II
IMP	imperative	INFIN	infinitive	INSTR	instrumental
INTERJ	interjection	IRR	irrealis	ITER	iterative
KOR	Koryak	LOC	locative	NDIR	non-directed (motion)
NEG	negative	NML	nominalizer	OBJ	object
PL	plural	PRES	present	PRON	pronoun
PRT	participle	SG	singular	SUBJ	subject
TR	transitive	VOC	vocative	#>#	subject > object (agreement)

Example sources K2.*n* refer to the texts collected by Jochelson following the text numbering format in Worth (1961); among the 1993–94 sources, AS is the text *Angaqe Sisike*.

References

- Abramovitz, Rafael. 2015. Morphologically conditioned restrictions on vowel distribution in Koryak. Handout from MIT Phonology Circle.
- Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition* 119–161.
- Asinovskij, Aleksandr S. & Aleksandr P. Volodin. 1987. The typology of vocalic structures of the word in Chukchi-Kamchatkan languages. In Tamaz Gamkrelidze (ed.), *Proceedings of the eleventh international congress of phonetic sciences*, 362–364. Tallinn: Academy of Sciences of the Estonian SSR.
- Baković, Eric. 2000. *Harmony, dominance and control*. New Brunswick, NJ: Rutgers University dissertation.
- Bobaljik, Jonathan David. 1998. Mostly predictable: Cyclicity and the distribution of schwa in itelmen. In Vida Samiian (ed.), *Proceedings of the western conference on linguistics (1996)*, vol. 9, 14–28. Fresno: CSU.

- Bobaljik, Jonathan David. 2006. Itelmen reduplication: Edge-in association and lexical stratification. *Journal of Linguistics* 42. 1–23.
- Bobaljik, Jonathan David & David Koester. 1999. The first Itelmen author? *SSILA Newsletter* 17(4). 5–6.
- Bogoras, Waldemar. 1917. *Koryak texts*. Leyden: E. J. Brill.
- Bogoras, Waldemar. 1922. Chukchee. In Franz Boas (ed.), *Handbook of american indian languages*, 631–903. Washington: Government Printing Office.
- Calabrese, Andrea. 1988. *Towards a theory of phonological alphabets*: MIT dissertation.
- Clements, George N. & Engin Sezer. 1982. Vowel and consonant disharmony in Turkish. In Harry van der Hulst & Norval Smith (eds.), *The structure of phonological representations*, 213–255. Dordrecht: Foris.
- Dras, Mark & K. David Harrison. 2003. Emergent behavior in phonological pattern change. In Russell K Standish, Mark A. Bedau & Hussein A. Abbass (eds.), *Artificial life*, vol. VIII, 390–393. MIT Press.
- Dresher, B. Elan & Aditi Lahiri. 2015. Romance loanwords and stress shift in English: A quantitative approach. Talk presented at the Second Edinburgh Symposium on Historical Phonology, University of Edinburgh, 3–4 December 2015.
- Dunn, Michael. 1999. *A grammar of Chukchi*: Australian National University dissertation.
- Fortescue, Michael. 1998. *Language relations across Bering Strait*. London: Cassell.
- Fortescue, Michael. 2005. *Comparative chukotko-kamchatkan dictionary*. Berlin: Mouton de Gruyter.
- Harrison, K. David, Mark Dras & Berk Kapicioglu. 2002. Agent-based modeling of the evolution of vowel harmony. In *Proceedings of NELS 32*, 217–236.
- Ito, Junko & Armin Mester. 1999. The phonological lexicon. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 62–100. Blackwell.
- Kabak, Barış. 2011. Turkish vowel harmony. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, 2831–2854. Oxford: Blackwell.
- Kenstowicz, Michael. 1979. Chukchee vowel harmony and epenthesis. In Paul R. Clyne, William F. Hanks & Carol L. Hofbauer (eds.), *Papers from “the elements,” a parasession*, 402–412. Chicago: Chicago Linguistics Society.
- Khaloimova, Klavdia N., Michael Dürr & Erich Kasten. 2014. *Ительменские сказки собранные В.И. Иохельсоном в 1910–1911 гг.* Fürstenberg, Germany: Kulturstiftung Sibirien.
- Kiparsky, Paul. 1973. Phonological representations. In Osamu Fujimura (ed.), *Three dimensions of linguistic theory*, 3–136. Tokyo: Taikusha.
- Koester, David & Jonathan David Bobaljik. 1994. Minority language, cultural revival and native rights in Russia: The Itel'men language as a case study. *NCSEER Working Papers*.
- Krause, Scott. 1979. *Topics in Chukchee phonology and morphology*. Urbana-Champaign, IL: University of Illinois dissertation.
- Kurebito, Tokusu. 2012. An outline of valency-reducing operations in Chukchi. In Wataru Nakamura & Ritsuko Kikusawa (eds.), *Objectivization and subjectivization: A typology of voice systems*, vol. 77 *Senri Ethnological Studies*, 177–189.
- Lightner, Theodore M. 1965. On the description of vowel and consonant harmony. *Word* 19. 376–387.
- Mel'nikov, G. I. 1948. Фонемы чукотского языка. *Язык и мышление* 11. 208–229.
- Moskal, Beata. 2015. *Domains on the border: between morphology and phonology*. Storrs, CT: University of Connecticut Ph.d. dissertation.

- Muravyova, Irina A. 1979. *Сопоставительное исследование морфонологии чукотского, корякского и алюторского языков*. Moscow: Moscow State University dissertation.
- Skorik, Piotr Ja. 1961. *Грамматика чукотского языка*, vol. 1. Moscow: Izdatel'stvo akademii nauk SSSR.
- Skorik, Piotr Ja. 1977. *Грамматика чукотского языка, часть II: глагол, наречие, служебные слова*. Leningrad: Nauka.
- Stebnickij, S. N. 1934. Ительменский язык. In G. N. Prokof'ev, E. A. Kreinovich & Ja. P. Al'kor (eds.), *Языки и письменность народов севера*, 85–104. Moscow: Gos. Uchpedgiz.
- Volodin, Aleksandr P. 1976. *Ительменский язык*. Moscow: Nauka.
- Volodin, Aleksandr P. 2003. *Ительмены*. Saint Petersburg: Drofa.
- Volodin, Aleksandr P. & Klavdia N. Khaloimova. 1989. *Словарь ительменско-русский и русско-ительменский*. Leningrad: Prosveshchenie.
- Worth, Dean S. 1961. *Kamchadal texts collected by W. Jochelson*. Berlin: Mouton.
- Worth, Dean S. 1969. *Dictionary of Western Kamchadal*. Berkeley: University of California Press.
- Yang, Charles. 2016. *The price of linguistic productivity*. Cambridge: MIT Press.