Proceedings of CLS 59 (2024), 319-330

Tone-sensitive vowel lengthening in BCMS monosyllables

Aljoša Milenković
Harvard University

1 Introduction

Bosnian/Croatian/Montenegrin/Serbian (BCMS) displays a typologically rare monosyllabic lengthening (ML) process which is crucially sensitive to tone. ML targets toneless (1a), but not High-toned monomoraic inputs (1b) (Zec 1999).

(1)	a.	ˈlɛːd	‘ice.NOM.SG’	ˈlɛda	‘ice.GEN.SG’
		ˈstɔː	‘hundred’	ˈpɛːtstɔ	‘five hundred’
		ˈbɔːs	‘barefoot.NOM.SG.M’	ˈbɔsi	‘barefoot.NOM.PL.M’
	b.	ˈdɛ́d	‘grandfather.NOM.SG’	ˈdɛ́da	‘grandfather.GEN.SG’
		ˈzdráʋ	‘healthy.NOM.SG.M’	ˈzdráʋi	‘healthy.NOM.PL.M’
		ˈlíx	‘pour.AOR.1SG’	ˈlísmɔ	‘pour.AOR.1PL’

The alternation in (1a) arises from vowel lengthening in monosyllabic forms rather than vowel shortening in polysyllables, as evident from the absence of such polysyllabic shortening process in the language (2).

(2)	ˈʋuːk	‘wolf.NOM.SG’	ˈʋuːka	‘wolf.GEN.SG’
	ˈgraːd	‘town.NOM.SG’	ˈgraːda	‘town.GEN.SG’
	ˈzuːb	‘tooth.NOM.SG’	ˈzuːba	‘tooth.GEN.SG’

In this paper, I discuss possible analyses and theoretical implications of the puzzling lengthening process in BCMS monosyllables. I argue that the process is driven jointly by prosodic minimality requirements and constraints governing tone-stress interaction. This complex prosodic pattern bears important implications for prosodic theory, and for the debate between different constraint-based phonological frameworks.

2 Issues

This section addresses two central questions raised by the lengthening pattern in (1): the issue of the driving force behind ML (§2.1), and of the source of the dual behavior of toneless and High-toned monosyllables with respect to ML (§2.2).

2.1 Diving force

Leaving aside for the time being the fact that ML is restricted to toneless monosyllables, I present two analyses of the process: the faithfulness-based account, which asserts that ML instantiates compensatory lengthening (CL), and the markedness-based account, which links ML to prosodic minimality requirements in BCMS. I will ultimately reject the CL analysis and pursue the prosodic minimality analysis.

The question of the driving force behind ML can be approached from a synchronic and diachronic perspective. It is important to draw a clear distinction between the synchronic and diachronic analyses of the process, since the historical source of the length alternation in (1a) need not necessarily coincide with the synchronic mechanism driving ML in BCMS toneless monosyllables. This paper aims primarily to shed light on the synchronic motivation for ML, and only discusses diachrony to the extent that it bears on the synchronic account.

The main rationale for the CL analysis of ML stems from diachronic considerations. Late Common Slavic (LCS), the linguistic predecessor of BCMS, featured a ban on closed syllables. Thus, the LCS preforms of consonant-final forms in (1) had a final lax vowel. There were two lax vowels in LCS: one front (*ĭ) and one back (*ŭ), which are known in Slavic linguistics as yer vowels. Final yers were lost historically, giving rise to hitherto intolerable closed syllables. The loss of final yers has in turn been thought to have led to compensatory lengthening in BCMS monosyllables that lack underlying prominence (Timberlake 1983; Hock 1986; Kavitskaya 2002); cf. (3).

(3)	LCS	BCMS	Gloss
	*ledŭ	lɛːd	‘ice.NOM.SG’
	*kostĭ	kɔːst	‘bone.NOM.SG’
	*bosŭ	bɔːs	‘barefoot.NOM.SG.M’

Assuming that final yers were not lost irrecoverably and are still part of the underlying representations of the forms in (1), it may be possible to extend the CL analysis to ML in modern BCMS. However, the CL analysis runs into theoretical and empirical problems. First, as Kavitskaya (2002) points out, cases of synchronically productive CL through vowel deletion are extremely rare because the CL-inducing vowels are often not synchronically transparent, leading to the reanalysis of resulting length as either underlying, or arising via some different lengthening mechanism. Furthermore, even if the length in toneless monosyllables in BCMS was historically brought about by CL, the CL analysis fails to explain the full range of ML environments in BCMS. Unequivocal evidence that not all vowels in BCMS toneless monosyllables lengthen compensatorily is provided by the patterning of the second and third person singular forms of the aorist. In AOR.2/3SG, which corresponds to the bare verbal stem, toneless monomoraic stems undergo ML. ML in AOR.2/3SG cannot be attributed to final yer deletion because no final yer can be posited underlyingly for this form, nor has there ever been a final yer in the form’s history, as indicated by the comparison of the BCMS forms with their LCS preforms in Table 1 (for the LCS aorist, see Vaillant 1966).

	LCS		BCMS	
	singluar	plural	singular	plural
1	*lixŭ	*lixomŭ	ˈlíx	ˈlísmɔ
2	*li	*liste	ˈliː	ˈlístɛ
3	*li	*lišę	ˈliː	ˈlíʃɛ

Table 1: The inflection of the aorist in LCS and BCMS

While the CL process /CVC_0ŭ/ → [$CV{:}C_o$] may seem appealing for (1a), it does evidently not carry over to AOR.2/3SG, where ML targets an originally monomoraic form: /li/ → [liː]. The reason why the AOR.2/3SG forms of some verbs are toneless, while the rest of the paradigm displays a High tone (as in Table 1) is the morphophonological rule that deletes lexical Highs in AOR.2/3SG. Whether an underlyingly High-toned verbal stem deletes its lexical High in AOR.2/3SG is unpredictable from the verb's phonological or any other properties. Consequently, there are two classes of High-toned verbal stems in BCMS (Daničić 1896): those that retain their lexical High in AOR.2/3SG and are not subject to ML (4a), and those that delete their High tone in AOR.2/3SG and undergo ML if monomoraic udnerlyingly (4b). Effectively, the tone deletion rule feeds ML in AOR.2/3SG of tone-deleting monomoraic verbal stems.

(4) a. *High-preserving verbs*

		INF	PTCP.F.SG	AOR.1PL	AOR.1SG	AOR.2/3SG
/slá-/	‘send’	ˈsláti	ˈslála	ˈslásmɔ	ˈsláx	ˈslá
/bí-/	‘beat’	ˈbíti	ˈbíla	ˈbísmɔ	ˈbíx	ˈbí
/ʧú-/	‘hear’	ˈʧúti	ˈʧúla	ˈʧúsmɔ	ˈʧúx	ˈʧú
/ʒɛ́-/	‘reap’	ˈʒɛ́ti	ˈʒɛ́la	ˈʒɛ́smɔ	ˈʒɛ́x	ˈʒɛ́
/tŕ̩-/	‘rub’	ˈtŕ̩ti	ˈtŕ̩la	ˈtŕ̩smɔ	ˈtŕ̩x	ˈtŕ̩

b. *High-deleting verbs*

		INF	PTCP.F.SG	AOR.1PL	AOR.1SG	AOR.2/3SG
/lí-/	‘pour’	ˈlíti	ˈlíla	ˈlísmɔ	ˈlíx	ˈliː
/ʃí-/	‘sew’	ˈʃíti	ˈʃíla	ˈʃísmɔ	ˈʃíx	ˈʃiː
/mí-/	‘wash’	ˈmíti	ˈmíla	ˈmísmɔ	ˈmíx	ˈmiː
/zná-/	‘know’	ˈznáti	ˈznála	ˈznásmɔ	ˈznáx	ˈznaː
/sjá-/	‘shine’	ˈsjáti	ˈsjála	ˈsjásmɔ	ˈsjáx	ˈsjaː

The upshot of the present discussion is that the CL analysis outlined herein does not provide a viable *synchronic* account of the lengthening facts in BCMS monosyllables. Whether CL was the historical source of vowel length in BCMS toneless monosyllables need not detain us much further. Note, however, that there are serious challenges to the analysis holding that this length was historically brought about via CL. In (3), a yer was lost in the final syllable, supposedly leading to the lengthening of the preceding vowel, which yields the expected right-to-left directionality in vowel loss-induced CL. However, lengthening is likewise observed in forms where a yer was lost in the syllable preceding the lengthened vowel, as in LCS **sŭto* → BCMS stɔː ‘hundred’. This poses a problem to the CL account because left-to-right CL through vowel loss is argued to be an impossible sound change (Hayes 1989; Kavitskaya 2002). Therefore, the loss of yer vowels need not have been the direct diachronic trigger of vowel lengthening in BCMS monosyllables. Rather, yer loss was relevant insofar as it introduced a previously non-existing class of monomoraic forms, which were eliminated by vowel lengthening to avoid the violation of the minimal foot constraint, on which I will elaborate in what follows.

I conclude that yer deletion is not the synchronic trigger of ML, and was likely not the historical source of vowel length in toneless monosyllables in BCMS. Following Zec (1999), the purpose of ML is to eliminate prosodically deficient monomoraic feet. On Zec's (1999) Optimality Theoretic (OT; Prince & Smolensky 1993/2004) account, ML takes place to comply with FOOTBINARITY, the markedness constraint that penalizes monomoraic feet. FTBIN outranks DEP-μ, which prohibits vowel lengthening:

(5)

li	FTBIN	DEP-μ
a. ˈ(li)	$*$!	
b. ☞ ˈ(lii)		$*$

Unlike the CL analysis, the prosodic minimality analysis straightforwardly accounts for ML in all relevant prosodic environments, including AOR.2/3SG. However, the bimoraic prosodic minimum is not strictly enforced in BCMS, a notable exception being High-toned monomoraic forms, which are exempt from ML (1b). The following section addresses the role of tone in defining the minimal size of prosodic constituents in BCMS.

2.2 Tonal conditioning

Having shown that ML has to do with the bimoraic prosodic minimum in BCMS, the question arises of why monomoraic forms are not prohibited across the board, in light of the fact that High-toned monomoraic forms are permitted in the language. There are two ways to explain why ML is restricted to toneless monosyllables. On one account, the dual prosodic behavior of toneless and High-toned monomoraic forms in BCMS instantiates a *blocking*, i.e. *do-something-except-when* effect (in the sense of Prince & Smolensky 1993/2004: §3–4), whereby a process generally applies except when it gives rise to an impermissible marked structure. The blocking analysis is pursued by Zec (1999), who provides the only synchronic generative account of ML to date. Zec holds that the relevant prosodic constraints are ranked such that vowel lengthening in monomoraic forms is the default in BCMS (cf. 5). ML is blocked in High-toned inputs because the resulting bimoraic trochee is disharmonic. Drawing on the observation that asymmetric heavy-light trochees are dispreferred cross-linguistically (Prince 1990), Zec argues that the ill-formedness of the ($\acute{\mu}\mu$) trochee results from the tonal asymmetry between its head mora, which is High-toned, and its nonhead mora, which is toneless. Per Zec, High-toned moras are metrically stronger than their toneless counterparts. Thus, the ($\acute{\mu}\mu$) trochee runs afoul of the preference for symmetric grouping in trochaic feet. The tonal contrast between the head and nonhead mora of a bimoraic trochee is penalized by the TROCHAICQUANTITY (TROCHQUANT) constraint. TROCHQUANT dominates FTBIN in BCMS, ensuring that lengthening is blocked in High-toned monosyllables:

(6)

slá	TROCHQUANT	FTBIN	DEP-μ
a. ☞ ˈ(slá)		$*$	
b. ˈ(sláa)	$*$!		$*$

Zec's blocking analysis successfully captures the problematic lengthening pattern in BCMS monosyllables. However, the account runs into several issues. First, the central theoretical assumption behind Zec's blocking analysis is that trochees with a higher-pitched first element are marked for the same reason asymmetric heavy-light trochees are considered disharmonic, implying that pitch-based grouping effects parallel duration-based grouping effects. However, experimental work on pitch-based grouping effects has found a trochaic bias in pitch-based rhythmic grouping (Bion & Nespor 2011; De la Mora *et al.* 2013; Crowhurst 2020), suggesting that pitch patterns with intensity rather than quantity with respect to rhythmic grouping, contrary to Zec (1999). Second, Zec's TROCHQUANT constarint produces a number of pathological effects. For instance, consider an OT grammar in which TROCHQUANT outranks MAX-μ. In this hypothetical language, toneless heavies surface faithfully under stress (7a), but heavy syllables with a High-toned first mora become light in stressed position (7b).

(7) a.

$\mu\mu$	TRQ	MAX-μ
a. ☞ $^{\text{ˈ}}(\mu\mu)$		
b. $^{\text{ˈ}}(\mu)$		$*!$

b.

$\acute{\mu}\mu$	TRQ	MAX-H	MAX-μ
a. $^{\text{ˈ}}(\acute{\mu}\mu)$	$*!$		
b. $^{\text{ˈ}}(\mu\mu)$		$*!$	
c. ☞ $^{\text{ˈ}}(\acute{\mu})$			$*$

Similarly, TROCHQUANT can lure stress away from High-toned heavies to some less prominent prosodic position, as in (8). In (8a), stress falls on a heavy syllable to satisfy the WEIGHT-TO-STRESS PRINCIPLE (WSP), which penalizes stressed light syllables. However, in (8b), stress falls on a light syllable in response to TROCHQUANT, to avoid a bimoraic trochee with a High-toned head mora.[1]

(8) a.

$\mu.\mu\mu$	TRQ	WSP
a. ☞ $\mu.^{\text{ˈ}}(\mu\mu)$		
b. $^{\text{ˈ}}(\mu).\mu\mu$		$*!$

b.

$\mu.\acute{\mu}\mu$	TRQ	WSP
a. $\mu.^{\text{ˈ}}(\acute{\mu}\mu)$	$*!$	
b. ☞ $^{\text{ˈ}}(\mu).\acute{\mu}\mu$		$*$

In sum, Zec's (1999) TROCHQUANT constraint treats bimoraic trochees headed by a High-toned mora as marked, although both bimoraic feet and High-toned foot heads are individually unmarked. Thus, Zec's account holds that the combination of two unmarked structures incurs a markedness penalty. More concretely, the TROCHQUANT constraint makes problematic typological predictions, as highlighted above.

A way around TROCHQUANT's problematic effects, while maintaining the gist of the blocking analysis, may be provided by a positional faithfulness account (Beckman 1998). In (9), I introduce a positional faithfulness constraint that prohibits vowel lengthening in High-toned syllables. Thus, the absence of ML in High-toned monosyllables falls out from the privileged status of High-toned syllables, which hinders certain unfaithful mappings in this environment. As in (5), ML arises from FTBIN's dominance over the context-free anti-lengthening constraint

[1] In (8), dots indicate syllable breaks.

DEP-μ. However, FTBIN ranks below the positional faithfulness constraint DEP-μ_H, which protects the underlying quantity of High-toned vowels. These rankings correctly restrict ML to toneless monosyllables, as shown in (9).

(9)

slá	DEP-μ_H	FTBIN	DEP-μ
a. ☞ ˈ(slá)		*	
b. ˈ(sláa)	*!		*

In (9), the positional faithfulness constraint DEP-μ_H achieves the same effect as Zec's TROCHQUANT. Furthermore, DEP-μ_H does so at no additional cost, as it does not share TROCHQUANT's flawed typological predictions. However, the positional faithfulness analysis is not devoid of problems. Beckman (1998) does not report any cases of positional privilege effects involving High-toned syllables. I am likewise not aware of any such positional faithfulness effects identified elsewhere in the literature. Positional faithfulness effects are usually associated with prosodic prominence (e.g. word-level stress), or a particular position within a morphosyntactic constituent (e.g. root- or stem-initial syllables). While tonal inventories can be positionally restricted such that more tonal contrasts occur in a more prominent environment, evidence for positional faithfulness effects treating High-toned syllables as the locus of enhanced faithfulness seems to be lacking. In fact, there is no independent evidence for the privileged status of High-toned syllables in BCMS, and hence no motivation for the positional faithfulness constraint in (9) other than the faithful realization of High-toned monosyllables.

In light of the deficiency of both blocking analyses, I pursue a different approach to the prosodic patterning of BCMS monosyllables, which ascribes this complex pattern to a *triggering*, i.e. *do-something-only-when* process interaction, whereby a mapping procedure is generally inoperative in the grammar, except to eliminate an impermissible marked structure. On the triggering analysis, FTBIN, the lengthening-inducing constraint, is overridden by faithfulness, which makes the absence of ML the default. The asymmetry between toneless and High-toned monosyllables does not arise because there is something problematic about bimoraic trochees headed by a High-toned mora, but rather because there is something problematic about monomoraic feet headed by a toneless mora, the structure eliminated by ML. This move shifts the analytical burden away from the issue of High-toned monosyllables' resistance to lengthening (which follows from DEP-$\mu \gg$ FTBIN) to the issue of the motivation for ML *specifically* in toneless monomoraic forms.

In BCMS, lexical Highs determine the position of word stress (Inkelas & Zec 1988; de Lacy 2002; Zec & Zsiga 2010). A curious property of BCMS which sets it apart from typical languages with tone-driven stress is that lexical Highs not only attract stress, but also join forces with prosodic weight in defining the minimal acceptable foot shape. Whether a foot is well-formed in BCMS depends not only on the number of moras it incorporates, but also on the tonal profile of its head mora. The present analysis draws on the following descriptive generalizations: (i) monomoraic feet are tolerated in BCMS iff their only mora bears a lexical High (cf. 1b); (ii) feet headed by a toneless mora are permitted iff they are bimoraic (cf. 2); conesequently (iii) monomoraic feet headed by a toneless mora are strictly prohibited in BCMS. I propose that ML is a repair strategy employed to eliminate toneless feet with a toneless head mora.

The illicitness of toneless, but not of High-toned monomoraic feet in BCMS indicates that the avoidance of monomoraic feet cannot be the sole driving force behind ML. Toneless degenerate feet violate two markedness constraints, namely FtBin and Head-H (Yip 2001), which requires that foot heads be associated with a High tone. Descriptively, BCMS tolerates both monomoraic feet and feet with a toneless head mora, which suggests that FtBin and Head-H are both individually overridden by faithfulness, as demonstrated in (10a–10b).

(10) a. FtBin *overridden by Faithfulness*

slá			Faith	FtBin
a.	☞	ˈ(slá)		*
b.		ˈ(sláa)	*!	

b. Head-H *overridden by Faithfulness*

graad			Faith	Head-H
a.	☞	ˈ(graad)		*
b.		ˈ(gráad)	*!	

However, the combination of the individually tolerable marked structures in (10)—monomoraic feet with a toneless head mora—is categorically prohibited in BCMS. While FtBin alone cannot instigate ML, given its ranking below Faithfulness (10a), ML takes place whenever FtBin and Head-H are violated simultaneously. This suggests that two markedness constraints that individually yield to faithfulness (FtBin and Head-H) cannot be violated simultaneously in BCMS. The coincident violation of FtBin and Head-H is avoided at the expense of a higher-ranking faithfulness constraint—Dep-μ. This constraint interaction instantiates a ganging-up cumulativity effect (Jäger & Rosenbach 2006; Pater 2009b; Albright 2012; Breiss 2020). Ganging-up arises when two constraints $\mathbb{B}$ and $\mathbb{C}$ which are individually overridden by some constraint $\mathbb{A}$ jointly prevail against $\mathbb{A}$ in order to avoid their coincident violation. In BCMS, FtBin and Head-H gang up against Dep-μ. As a result, doubly-marked toneless degenerate feet are prohibited. The fact that individual violations of FtBin and Head-H are tolerable in BCMS, but their coincident violations are not, readily explains why ML is restricted to toneless monosyllables: the process does not eliminate degenerate feet across the board, but those degenerate feet that incur a Head-H penalty.

Ganging-up cumulativity is famously problematic for strict-ranking OT. Because Dep-μ dominates FtBin (cf. 10a), the constraint grammar in (11) incorrectly outputs the faithful parse for toneless monomoraic inputs. I work out the specifics of the constraint-based account of this gang effect in §3.

(11)

li			Dep-μ	FtBin	Head-H
a.	💣	ˈ(li)		*	*
b.	☹	ˈ(lii)	*!		*

To summarize the points made in this section. ML takes place in compliance with the bimoraic prosodic minimum in BCMS. The most perspicuous way of handling the dual behavior of toneless and High-toned monomoraic forms in BCMS is provided by the triggering analysis advanced herein, which holds that ML is generally inhibited, unless to avoid doubly-marked toneless degenerate feet. The ban on toneless degenerate feet in BCMS instantiates a gang effect, since the simultaneous violation of two individually violable markedness constraints is avoided at the expense of a higher-ranking faithfulness constraint.

3 ML as a superlinear gang effect: analysis and implications

Without nonstandard devices, OT is incapable of modeling ganging-up constraint cumulativity (cf. 11). OT accommodates constraint ganging using local constraint conjunction (LCC; Smolensky 1993, 2006), which is defined as follows:

(12) LOCAL CONSTRAINT CONJUNCTION (LCC)
The local conjunction of two constraints $\mathbb{A}$ and $\mathbb{B}$, $\mathbb{A}\&_D\mathbb{B}$, assesses a mark whenever both $\mathbb{A}$ and $\mathbb{B}$ are violated within the designated local domain *D*.

To derive the lengthening facts in BCMS monosyllables, in particular the fact that toneless degenerate feet are intolerable, while their High-toned counterparts are permitted, I introduce the local conjunction of FTBIN and HEAD-H.[2] FTBIN&HEAD-H outranks the faithfulness constraint DEP-μ, shown in §2.2 to rank above FTBIN. These rankings correctly enforce vowel lengthening in toneless, but not in High-toned monomoraic input forms. The LCC analysis of the prosodic behavior of BCMS monomoraic forms is provided in (13).

(13) a.

slá	FTB&HD-H	DEP-μ	FTB	HD-H
a. ☞ ˈ(slá)			*	
b. ˈ(sláa)		*!		

b.

li	FTB&HD-H	DEP-μ	FTB	HD-H
a. ˈ(li)	*!		*	*
b. ☞ ˈ(lii)		*		*

In (13a), there is no coincident violation of FTBIN and HEAD-H, given that HEAD-H's demands are met. FTBIN&HEAD-H is inactive and the faithful parse is selected as optimal because it performs better on DEP-μ than the lengthening candidate. In (13b), the faithful parse *[ˈ(li)] violates *both* FTBIN and HEAD-H, thereby incurring a mark to top-ranking FTBIN&HEAD-H. The ML candidate [ˈ(lii)] satisfies FTBIN&HEAD-H at the expense of DEP-μ.

OT enriched with LCC successfully models ganging-up cumulativity, as well as certain other process interactions, including counterfeeding opacity (cf. Moreton

[2]It is safe to assume that the domain of the conjoined constraint FTBIN&HEAD-H is the Prosodic Word (ω), which is the domain of stress assignment in most BCMS dialects.

& Smolensky 2002 and references therein). However, it has been pointed out that LCC grammars are overly powerful (Kirchner 1996; Padgett 2002; McCarthy 2003; Pater 2009a).

The standard version of Harmonic Grammar (HG; Legendre *et al.* 1990), which employs numeric weights rather than strict ranking to assess the severity of constraint violations, has been argued to be capable of deriving gang effects with no appeal to LCC (Potts *et al.* 2010; Farris-Trimble 2008; Pater 2009b; Pater 2016). To illustrate this, consider the hypothetical case of ganging-up cumulativity in (14), where two lower-weighted constraints $\mathbb{B}$ and $\mathbb{C}$ jointly outpower higher-weighted $\mathbb{A}$.

(14)

	$\mathscr{H}$	$\mathbb{A}$ 2	$\mathbb{B}$ 1.5	$\mathbb{C}$ 1.5
a. ☞ candidate$_1$	-2	-1		
b. candidate$_2$	-3		-1	-1

Given the propensity of LCC grammars to overgeneration, standard HG's ability to handle ganging-up cumulativity without LCC has been viewed as a major advantage of HG over OT. However, HG makes fairly strict predictions about possible gang effects. HG predicts that for a gang effect to be possible, there must be an asymmetric trade-off whereby a stronger constraint is sacrificed to comply with *multiple* weaker constraints (Pater 2009b, 2016), as in the hypothetical example in (14). Crucially, no gang effect can take place unless both ganging constraints are satisfied (Pater 2009b; Smith 2022). The concept of asymmetric trade-off is formalized in (15) ($w(x)$ = weight of x).

(15) *Asymmetric trade-off*

a. $w(\mathbb{A}) > w(\mathbb{B})$

b. $w(\mathbb{A}) > w(\mathbb{C})$

c. $w(\mathbb{B}\&\mathbb{C}) > w(\mathbb{A})$

The LCC approach to ganging-up cumulativity makes more flexible predictions. Gang effects can arise regardless of whether only one or both ganging constraints are satisfied: the local conjunction of the ganging constraints is satisfied as long as there is no coincident violation.

Consider now the gang effect of FTBIN and HEAD-H against DEP-μ in BCMS. In (13b), vowel lengthening is employed to remove the coincident violation of FTBIN and HEAD-H. This strategy satisfies FTBIN, while the winning candidate still violates HEAD-H.[3] The LCC account in (13) correctly derives ML in toneless monomoraic inputs. By contrast, an HG account of ML faces the same issue as the OT account without LCC (11): it erroneously favors the faithful parse for toneless monomoraic inputs, even when the constraints' weights are set to allow FTBIN and

[3] In order to satisfy both ganging constraints, two unfaithful mappings would have to take place: vowel lengthening and High tone insertion. Tone insertion is blocked by high-ranking DEP-H, which outranks HEAD-H.

HEAD-H to jointly prevail against DEP-μ. This is because the intended winner and its faithful contender share the violation of HEAD-H, as shown in (16).[4]

(16)

li		$\mathscr{H}$	DEP-μ 2	FTB 1.5	HD-H 1.5
a. 💣	ˈ(li)	-3		-1	-1
b. ☹	ˈ(lii)	-3.5	-1		-1

In BCMS, when FTBIN and HEAD-H are violated together, FTBIN is satisfied at the expense of a higher-weighted constraint—DEP-μ. This constraint interaction involves a symmetric trade-off between constraint violations (Pater 2009b; Pater 2016; Smith 2022), where a higher-weighed penalty (DEP-μ) is traded against a lower-weighted penalty (FTBIN) in HEAD-H-violating environments. Constraint ganging via symmetric trade-off is formalized in (17).

(17) *Symmetric trade-off*

a. $w(\mathbb{A}) > w(\mathbb{B})$

b. $w(\mathbb{B}\&\mathbb{C}) > w(\mathbb{A}\&\mathbb{C})$

Symmetric trade-offs are inherently superlinear (in the sense of Smith & Pater 2020; Breiss & Albright 2022), since for (17b) to hold, the cumulative weight of the ganging constraints must be greater than the linear sum of their individual weights. I subscribe the approach that employs weighted constraint conjunction as a means of modeling superlinear gang effects in HG (Green & Davis 2014; Shih 2017). To capture the superlinear cumulative interaction of FTBIN and HEAD-H in BCMS, I propose the weighted conjunction of these two constraints. The analysis is outlined in (18).

(18)

li		$\mathscr{H}$	DEP-μ 2	FTB 1.5	HD-H 1.5	FTB&HD-H 1
a.	ˈ(li)	-4		-1	-1	-1
b. ☞	ˈ(lii)	-3.5	-1		-1	

The crucial takeaway of the present analysis is that both OT and HG necessitate LCC to model gang effects featuring a symmetric trade-off. Thus, HG does not entirely dispense with LCC, as occasionally suggested in the HG literature. It should be noted that there is a difference between OT and HG in the way these frameworks use LCC to model gang effects. In OT, LCC is needed for both symmetric and asymmetric trade-offs. By contrast, HG necessitates LCC solely for symmetric trade-offs, while asymmetric trade-offs are derivable without LCC (cf. 14).

In sum, I demonstrated that the gang effect identified in §2 involves a symmetric trade-off between a higher-weighted faithfulness constraint (DEP-μ) and a

[4] Recall that DEP-μ is known independently from the absence of lengthening in High-toned monosyllables to outweigh FTBIN.

lower-weighted markedness constraint (FTBIN). To model this kind of constraint cumulativity, both OT and HG necessitate local constraint conjunction. Future work must address the typological consequences of LCC in both OT and HG.

4 Conclusion

In this paper, I presented a tone-sensitive vowel lengthening process in BCMS monosyllables. I rejected the analysis that views this lengthening process as compensatory and argued that the process is driven by the bimoraic prosodic minimum in BCMS. I demonstrated that the triggering analysis of the process is superior to the blocking analyses. The paper contributes to prosodic typology by identifying a complex interplay between tone, syllable weight and metrical structure in BCMS whereby tone not only determines the locus of stress but also plays a role in defining the minimal size of prosodic constituents in the language. This, to my knowledge, is the first case of tone-sensitive prosodic minimality reported in the literature. Another theoretical finding of the study is the documentation of a gang effect that cannot be modeled in the HG framework without nonstandard mechanisms. The identification of such gang effect further corroborates the necessity of local constraint conjunction in both OT and HG.

References

Albright, A., 2012. Additive markedness interactions in phonology. Unpublished manuscript, MIT.

Beckman, J. N. 1998. *Positional faithfulness*. University of Massachusetts Amherst dissertation.

Bion, Ricardoand Benavides-Varela, S., & M. Nespor. 2011. Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and speech* 54. 123–140.

Breiss, C. 2020. Constraint cumulativity in phonotactics: Evidence from artificial grammar learning studies. *Phonology* 37. 551–576.

Breiss, C., & A. Albright. 2022. Cumulative markedness effects and (non-) linearity in phonotactics. *Glossa: a journal of general linguistics* 7.

Crowhurst, M. 2020. The iambic/trochaic law: Nature or nurture? *Language and Linguistics Compass* 14.

Daničić, Đ. 1896. *Akcenti u glagola*. Zagreb: Jugoslavenska akademija znanosti i umjetnosti.

De la Mora, D., M. Nespor, & J. Toro. 2013. Do humans and nonhuman animals share the grouping principles of the iambic–trochaic law? *Attention, Perception, & Psychophysics* 75. 92–100.

de Lacy, P. 2002. The interaction of tone and stress in Optimality Theory. *Phonology* 19. 1–32.

Farris-Trimble, A. 2008. *Cumulative faithfulness effects in phonology*. Indiana University dissertation.

Green, C., & S. Davis. 2014. Superadditivity and limitations on syllable complexity in bambara words. In *Perspectives on phonological theory and development, in honor of Daniel A. Dinnsen*, ed. by A. Farris-Trimble & J. Barlow, 223–247. Amsterdam: John Benjamins Publishing Company.

Hayes, B. 1989. Compensatory lengthening in moraic phonology. *Linguistic inquiry* 20. 253–306.

Hock, H. H. 1986. Compensatory lengthening: in defense of the concept 'mora'. *Folia Linguistica* 16. 431–460.

Inkelas, S., & D. Zec. 1988. Serbo-Croatian pitch accent: the interaction of tone, stress, and intonation. *Language* 64. 227–248.

Jäger, G., & A. Rosenbach. 2006. The winner takes it all—almost: Cumulativity in grammatical variation. *Linguistics* 44. 937–971.

Kavitskaya, D. 2002. *Compensatory lengthening: phonetics, phonology, diachrony*. Routledge.

Kirchner, R. 1996. Synchronic chain shifts in optimality theory. *Linguistic inquiry* 27. 341–350.

Legendre, G., Y. Miyata, & P. Smolensky. 1990. Harmonic grammar—a formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. In *Proceedings of the twelfth annual conference of the Cognitive Science Society*, 884–891. Citeseer.

McCarthy, J. 2003. Comparative markedness. *Theoretical Linguistics* 29. 1–51.

Moreton, E., & P. Smolensky. 2002. Typological consequences of local constraint conjunction. In *Proceedings of the Twenty-First West Coast Conference on Formal Linguistics*, ed. by L. Mikkelsen & C. Potts, 306–319. Cascadilla Press.

Padgett, J., 2002. Constraint conjunction versus grounded constraint subhierarchies in optimality theory. Unpublished manuscript, University of California Santa Cruz.

Pater, J. 2009a. [Review of] Paul Smolensky and Géraldine Legendre (2006). The harmonic mind: from neural computation to optimality-theoretic grammar. Cambridge, Mass.: MIT Press. Vol. 1: Cognitive architecture. Pp. xxiv+ 563. Vol. 2: Linguistic and philosophical implications. Pp. xxiv+ 611. *Phonology* 26. 217–226.

Pater, J. 2009b. Weighted constraints in generative linguistics. *Cognitive science* 33. 999–1035.

Pater, J. 2016. Universal grammar with weighted constraints. In *Harmonic grammar and harmonic serialism*, ed. by J. Pater & J. McCarthy, 1–46. London: Equinox.

Potts, C., J. Pater, K. Jesney, R. Bhatt, & M. Becker. 2010. Harmonic grammar with linear programming: from linear systems to linguistic typology. *Phonology* 27. 77–117.

Prince, A. 1990. Quantitative consequences of rhythmic organization. In *CLS 26: Volume 2: The Parasession*, ed. by M. Ziolkowski, M. Noske, & K. Deaton, 355–398. Chicago Linguistic Society.

Prince, A., & P. Smolensky. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA: Blackwell. Technical Report, Rutgers University and University of Colorado at Boulder, 1993. Revised version Blackwell, 2004.

Shih, S. 2017. Constraint conjunction in weighted probabilistic grammar. *Phonology* 34. 243–268.

Smith, B., & J. Pater. 2020. French schwa and gradient cumulativity. *Glossa: a journal of general linguistics* 5.

Smith, J. 2022. Some formal implications of deletion saltation. *Linguistic inquiry* 53. 852–864.

Smolensky, P., 1993. Harmony, markedness, and phonological activity. Paper presented at Rutgers Optimality Workshop.

Smolensky, P. 2006. Optimality in phonology II: harmonic completeness, local constraint conjunction, and feature domain markedness. In *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar. Vol. 2: Linguistic and pholosophical implications*, ed. by P. Smolensky & G. Legendre, 27–160. Cambridge, MA: MIT Press.

Timberlake, A. 1983. Compensatory lengthening in slavic, 2: Phonetic reconstruction. In *American contributions to the Ninth International Congress of Slavists : Kiev, September 1983*.

Vaillant, A. 1966. *Grammaire comparée des langues slaves: Le verbe. Deuxième partie*. Paris: Éditions Klincksieck.

Yip, M. 2001. The complex interaction of tone and prominence. In *NELS*, volume 31, 531–545.

Zec, D. 1999. Footed tones and tonal feet: rhythmic constituency in a pitch-accent language. *Phonology* 16. 225–264.

Zec, D., & E. Zsiga. 2010. Interaction of tone and stress in standard serbian: Phonological and phonetic evidence. In *FASL 18: The Cornell Meeting 2008*, ed. by W. Browne, A. Cooper, A. Fisher, E. Kesici, N. Predolac, & D. Zec, 536–555. Michigan Slavic Publications.