

ARTICLE

Superadditive cumulativity in categorical prosodic patterns: Prosodic minimality in Bosnian/Croatian/Montenegrin/Serbian

Aljoša Milenković<sup>1</sup> 

<sup>1</sup>Department of Linguistics, Harvard University, Cambridge, MA, USA  
Email: [aljosamilenkovic@g.harvard.edu](mailto:aljosamilenkovic@g.harvard.edu).

Received: 23 September 2023; Revised: 09 June 2024; Accepted: 05 January 2025

**Keywords:** Tone, Stress, Prosodic Minimality, Cumulativity, Superadditivity, Local Constraint Conjunction, Optimality Theory, Harmonic Grammar

Abstract

This paper presents a case of SUPERADDITIVE GANGING-UP CUMULATIVITY in the metrical phonology of Bosnian/Croatian/Motenegrin/Serbian (BCMS). BCMS individually permits monomoraic feet and feet with a toneless head mora, but prohibits toneless monomoraic feet. Across BCMS dialects, several prosodic processes conspire against this doubly-marked structure. Due to the superadditive character of this interaction, both Optimality Theory and, importantly, Harmonic Grammar necessitate LOCAL CONSTRAINT CONJUNCTION to capture the ban on toneless monomoraic feet in BCMS. This demonstration constitutes evidence for conjoined constraints in weighted constraint grammar. The study contributes to the typology of cumulativity effects by documenting superadditive ganging in a categorical prosodic pattern, with virtually all previously reported cases of superadditivity being observed in variable phonological patterns.

Contents

1. Introduction	2
2. Prosodic minimality in BCMS	5
2.1 Background	5
2.1.1 Neoštokavian	6
2.1.2 Old Štokavian	7
2.1.3 Interim summary	10
2.2 The minimal foot in BCMS	11
2.3 Monosyllabic lengthening	11
2.4 Tonal Flop and Penultimate Lengthening	13
2.5 Conspiracy	15
3. Proposal	16
3.1 Joint and independent activity of HEAD-H and FtBIN in BCMS	16

3.2 OT analysis . . . . . 18

    3.2.1 Monosyllabic Lengthening via local constraint conjunction . . 18

    3.2.2 Tonal Flop and Penultimate Lengthening: constraint-based analysis . . . . . 20

3.3 HG analysis . . . . . 23

    3.3.1 Prosodic minimality in BCMS in standard HG . . . . . 24

    3.3.2 Weighted constraint conjunction in HG . . . . . 27

3.4 No adverse effects of constraint conjunction . . . . . 30

4. Alternatives . . . . . 31

    4.1 No uneven trochees: Zec 1999 . . . . . 31

    4.2 Theoretical and empirical issues with TrochaicQuantity . . . . . 33

        4.2.1 TrochaicQuantity and pitch-based rhythmic grouping . . . . . 33

        4.2.2 TrochaicQuantity penalises doubly-unmarked structure . . . . . 33

        4.2.3 Undergeneration . . . . . 35

    4.3 Section summary . . . . . 37

5. Conclusion . . . . . 38

1. Introduction


Constraint-based theories of phonology employ two strategies to assess the severity of constraint violation: strict ranking and numeric weights. In strict-ranking Optimality Theory (OT; Prince & Smolensky, 2004), the winner is the candidate that fares best on the highest-ranking constraint(s), irrespective of the overall violation profile. In Harmonic Grammar (HG; Legendre et al., 1990), well-formedness is computed by the weighted sum of all constraint violations.

HG predicts that the combined effect of multiple lower-weighted penalties can surpass the effect of violating a single higher-weighted constraint. Constraint cumulativity effects are of two types: GANGING-UP and COUNTING CUMULATIVITY (Jäger & Rosenbach, 2006). GANGING, exemplified in the HG tableau in (1), arises when violating constraints B and C individually is less severe than violating A, but violating B and C together surpasses violating A alone (Pater, 2009b; Kenstowicz, 2009; Farris-Trimble, 2008; Albright, 2012; McPherson, 2016; Ryan, 2017; Breiss, 2020; Breiss & Albright, 2022). In COUNTING CUMULATIVITY (2), multiple violations of a weaker constraint B incur a more severe penalty than a single violation of higher-weighted A (Kawahara, 2020; Breiss, 2020; Kawahara & Kumagai, 2021; Kawahara & Breiss, 2021; Kim, 2022).

(1) Ganging-up cumulativity

			A	B	C
$\mathcal{H}$			2	1.5	1.5
a.	✗	cand1	-2	-1	
b.		cand2	-3	-1	-1

(2) Counting cumulativity

		$\mathcal{H}$	A	B
			2	1.5
a.	 cand1	−2	−1	
b.	cand2	−3		−2

Constraint cumulativity has figured prominently in the debate between OT and HG. OT uses LOCAL CONSTRAINT CONJUNCTION (LCC; Smolensky, 1993, 2006) to capture ganging-up cumulativity. HG accommodates constraint cumulativity without LCC (cf. (1)–(2); Potts et al., 2010; Farris-Trimble, 2008; Pater, 2009b, 2016). This has been viewed as a major argument in favour of HG over OT (Pater, 2009a,b, 2016), in light of conjoined constraints’ propensity for overgeneration (cf. Kirchner, 1997; McCarthy, 2003, a.o.).

However, without nondefault mechanisms such as LCC, HG successfully models only a subset of cumulativity effects. Recent work (e.g. Smith & Pater, 2020; Breiss & Albright, 2022) distinguishes between three types of cumulative effects: LINEAR (i.e. ADDITIVE), SUPERLINEAR (i.e. SUPERADDITIVE) and SUBLINEAR (i.e. SUBADDITIVE). LINEARITY (3a) occurs when the effect on acceptability/probability of simultaneously violating constraints A and B equals the combination of independent effects of A and B. The combined effect of A and B is SUPERLINEAR if the effect of coincident violation of A and B exceeds the joint effect of their independent violations (3b). In SUBLINEAR cumulativity (3c), the effect of coincident violation of A and B results in a smaller decline in acceptability/probability than the combination of their independent effects.

- (3)
- a.

Linearity/Additivity

$$e(A) + e(B) = e(A, B)$$
- b.

Superlinearity/Superadditivity

$$e(A) + e(B) < e(A, B)$$
- c.


Sublinearity/Subadditivity

$$e(A) + e(B) > e(A, B)$$



where  $e(C)$  is the effect of violating constraint C and  $(C_1, C_2)$  is the coincident violation of  $C_1$  and  $C_2$

HG captures additive cumulativity with no appeal to nondefault devices (4), but not superadditive ganging-up cumulativity (Albright, 2009; Green & Davis, 2014; Shih, 2017; Smith, 2022). This is illustrated in (5), where candidate 1, the intended winner, loses out to candidate 2. (Suppose (i) there is independent evidence that A individually outweighs both B and C and (ii) the joint violation of B and C incurs a greater penalty than the joint violation of A and B.)

## (4) Additivity


		A	B	C
	$\mathcal{H}$	2	1.5	1.5
a.	 cand1	-2	-1	
b.	cand2	-3		-1

## (5) Superadditivity

		A	B	C
	$\mathcal{H}$	2	1.5	1.5
a.	 cand1	-3.5	-1	-1
b.	 cand2	-3		-1

To model superadditive constraint cumulatity, HG adopts weighted constraint conjunction (Albright, 2009; Green & Davis, 2014; Shih, 2017). In (5), the local conjunction of constraints **B** and **C** adds weight to their simultaneous violation, enabling candidate 1 to prevail against candidate 2 despite the shared violation of **B** (6). Unlike in OT (Itô & Mester, 1998; Baković, 2000), the conjoined constraint need not dominate individual conjuncts in HG. This follows from the cumulatity of constraint violations.

## (6)

		A	B	C	B&C
	$\mathcal{H}$	2	1.5	1.5	1
a.	 cand1	-3.5	-1	-1	
b.	cand2	-4		-1	-1

Thus, HG does not entirely dispense with LCC. However, the two frameworks make different use of LCC: OT employs LCC to model both additive and superadditive ganging, while HG utilises this device to model superadditivity.

This paper provides further evidence for the necessity of conjoined constraints in weighted constraint grammar through the examination of a complex prosodic minimality effect in Bosnian/Croatian/Montenegrin/Serbian (BCMS). BCMS tolerates both monomoraic feet and feet headed by a toneless mora alone, but not their combination—monomoraic feet with a toneless head mora. The prohibition on toneless monomoraic feet manifests itself in several processes, including tone-sensitive monosyllabic lengthening. Lengthening applies to toneless monomoraic forms (7a), but not to their High-toned counterparts (7b).

- (7) a. /lɛd/            ['lɛɛd]        'ice.NOM.SG'  
       b. /brát/        ['(brát)]      'brother.NOM.SG'        not \*['(bráat)]

This pattern illustrates the joint effect of two independently active constraints in the language: HEAD-H (Yip, 2001), which requires that foot-heading moras be High-toned, and FOOTBINARITY (Prince & Smolensky, 2004), which dictates that feet be binary at some level. I show that the coincident violation of HEAD-H and FOOTBINARITY results

in a greater decline in metrical well-formedness than expected from their independent effects, indicative of superlinearity.

This paper makes two contributions to the study of constraint cumulativity. First, contrary to the common position in the HG literature (Potts et al., 2010; Pater, 2009b), I show that additive cumulativity of simplex weighted constraints does not supplant the full range of effects of LCC (Green & Davis, 2014; Shih, 2017; Smith, 2022). This adds to the growing evidence for conjoined constraints in weighted constraint grammar (Albright, 2009; Hayes et al., 2012; Green & Davis, 2014; Shih, 2017). Second, the paper identifies superadditive cumulativity in a categorical prosodic pattern. Previous studies mainly examined superadditivity in variable patterns (Shih, 2017; Smith & Pater, 2020; Breiss & Albright, 2022; Kim, 2022).

Additionally, the paper identifies an interplay between prosodic minimality and tone in BCMS, contributing to prosodic typology. Tone is known to interact with stress (de Lacy, 2002; Gordon, 2023), and the BCMS patterns discussed herein indicate that tone also interacts with prosodic size constraints.

The paper is organised as follows. In §2, I present a perplexing tone-sensitive prosodic minimality effect in BCMS and provide new dialectal data that substantiate this effect. In §3, I propose a unified analysis of the patterns outlined in §2 and discuss their theoretical implications. §4 compares the present analysis with the alternative account proposed by Zec (1999), showing that only the present analysis captures all prosodic minimality phenomena in BCMS. §5 concludes.

## 2. Prosodic minimality in BCMS

### 2.1. Background

All BCMS dialects examined in this paper have lexical tone. Morphological structure and syllable weight are the primary factors that play a role in lexical tone assignment (Zec, 1999). Animacy predicts tonal class affiliation in some inflectional classes (Martinović, 2012). The crucial properties of the lexical tone system of BCMS are the following:

1. *Mora as the tone-bearing unit (TBU)*: Following Inkelas & Zec (1988), it is the mora rather than the syllable that functions as the TBU in BCMS (contra Langston, 1997; see Zsiga & Zec, 2013: 101 for arguments).
2. *Privativity and underspecification*: Underlyingly, moras are either High-toned or lack tone. Only High tone is specified in the lexicon; all moras that do not bear a High underlyingly are toneless and surface with a low pitch by default (Inkelas & Zec, 1988; Zec, 1999).
3. *Culminativity*: BCMS displays tonal culminativity, since the language allows for at most one singly-linked High tone per word. BCMS therefore falls into the category of restricted tone systems (Hyman, 2006).
4. *No rising contours*: Rising contour tones are strictly prohibited in Štokavian BCMS: Highs must be attached to the first mora of a heavy syllable (Ivić, 1958; Inkelas & Zec, 1988; Bethin, 1998).

In BCMS, lexical Highs determine the locus of stress (Inkelas & Zec, 1988; Zec, 1999; de Lacy, 2002; Zec & Zsiga, 2022). The two major dialect groups—Neoštokavian (NS) and Old Štokavian (OS)—share the same tonal system but differ in stress placement relative to High tone. This section provides a brief overview of the stress pattern of NS varieties of BCMS, and a more detailed overview of OS word prosody.

The word prosody of NS, the standard variety of BCMS, has been a topic of considerable debate. Extensive research has been conducted on the acoustics of NS stress (Lehiste & Ivić, 1986; Zsiga & Zec, 2013; Batas, 2014), as well as its phonological representation (Inkelas & Zec, 1988; Bethin, 1998; Zec & Zsiga, 2010). OS prosody has received comparably less attention in the generative literature than the prosody of NS (a notable exception being Zec & Zsiga, 2022). There are two primary reasons for this disparity. First, OS exhibits a high degree of prosodic variation across individual dialects (cf. Ivić, 1958), which makes it difficult to fit into a coherent theory. Second, most existing descriptive studies of OS are exclusively available in BCMS, limiting their accessibility to a broader audience.

### 2.1.1. Neoštokavian

In NS, lexical Highs determine stress position. Stress falls on the syllable immediately preceding the word's only High (8), or on the High-toned syllable if no syllable precedes (9). In underlyingly toneless inputs, NS displays initial stress with High tone insertion (10).

- |      |    |                    |                     |                      |
|------|----|--------------------|---------------------|----------------------|
| (8)  | a. | /u.sta.nɔ.ví.ti/   | [u.sta.'nɔ.ví.ti]   | 'establish.INF'      |
|      | b. | /u.ráa.diim/       | ['u.ráa.diim]       | 'do.PRS.1SG'         |
|      | c. | /dɔ.pri.nó.sii.mɔ/ | [dɔ.'pri.nó.sii.mɔ] | 'contribute.PRS.1PL' |
| (9)  | a. | /ráa.dii.mɔ/       | ['ráa.dii.mɔ]       | 'work.PRS.1PL'       |
|      | b. | /nó.sii.mɔ/        | ['nó.sii.mɔ]        | 'carry.PRS.1PL'      |
|      | c. | /dá.dɔ.fɛ/         | ['dá.dɔ.fɛ]         | 'give.AOR.3PL'       |
| (10) | a. | /prɔ.da.dɛ/        | ['prɔ.da.dɛ]        | 'sell.AOR.2/3SG'     |
|      | b. | /u.kra.dɛ/         | ['ú.kra.dɛ]         | 'steal.AOR.2/3SG'    |
|      | c. | /prɔ.pa.dɛ/        | ['prɔ.pa.dɛ]        | 'fail.AOR.2/3SG'     |

Forms in (9) and (10) are indistinguishable, with the tonal patterning of prefixes indicating whether the initial syllable is underlyingly High-toned (9) or has an inserted High (10). Forms with an initial lexical High retain their tone with a prefix (11), while underlyingly toneless forms have a High inserted on the prefix (12).

- |      |                           |                 |                  |                       |
|------|---------------------------|-----------------|------------------|-----------------------|
| (11) | Stem-initial lexical High |                 |                  |                       |
|      | a.                        | /nɛ-ráa.dii.mɔ/ | ['nɛ.ráa.dii.mɔ] | 'not-work.PRS.1PL'    |
|      | b.                        | /nɛ-nó.sii.mɔ/  | ['nɛ.nó.sii.mɔ]  | 'not-carry.PRS.1PL'   |
|      | c.                        | /nɛ-dá.dɔ.fɛ/   | ['nɛ.dá.dɔ.fɛ]   | 'not-give.AOR.3PL'    |
| (12) | No lexical High           |                 |                  |                       |
|      | a.                        | /nɛ-prɔ.da.dɛ/  | ['nɛ.prɔ.da.dɛ]  | 'not-sell.AOR.2/3SG'  |
|      | b.                        | /nɛ-u.kra.dɛ/   | ['nɛ.u.kra.dɛ]   | 'not-steal.AOR.2/3SG' |
|      | c.                        | /nɛ-prɔ.pa.dɛ/  | ['nɛ.prɔ.pa.dɛ]  | 'not-fail.AOR.2/3SG'  |

Bethin (1994, 1998) analyses NS as a SYLLABIC TROCHEE system.<sup>1</sup> This analysis sheds light on stress-tone adjacency in NS (8): the stressed and High-toned syllable must be contained within a disyllabic trochee, as shown in (13).

- (13) NS syllabic trochee  
 /u.ráa.diim/ → [(‘u.ráa).diim] ‘do.PRS.1SG’

Thus, NS stress assignment is QUANTITY-INSENSITIVE: stress falls on the syllable preceding a lexical High regardless of syllable weight (8).

### 2.1.2. Old Štokavian

As a general tendency, stress in OS typically falls on the High-toned syllable (Zec & Zsiga, 2022), as shown in (14).

- (14)
- |    |                   |                    |                |
|----|-------------------|--------------------|----------------|
| a. | /sɛ.kí.ra/        | [sɛ.‘kí.ra]        | ‘axe.NOM.SG’   |
| b. | /ka.púut/         | [ka.‘púut]         | ‘coat.NOM.SG’  |
| c. | /u.ráa.dii.mɔ/    | [u.‘ráa.di.mɔ]     | ‘do.PRS.1SG’   |
| d. | /prɔ.dáav.ni.tsa/ | [prɔ.‘dáav.ni.tsa] | ‘store.NOM.SG’ |

The preference for stress on High-toned syllables arises from the HEAD-H constraint (Yip, 2001), defined in (15).

- (15) HEAD-H  
 Every foot-heading mora must bear a High tone.

Despite the general preference for stressed High-toned syllables in OS, stress does not invariably fall on the High-toned syllable in all prosodic environments in all OS dialects (Zec & Zsiga, 2022).

There are two cases in which some OS dialects avoid stressing the High-toned syllable. The first point of variation across individual OS dialects stems from the prosodic patterning of final light syllables. When a lexical High is attached to a final light syllable, stress falls on the final light in only a handful of OS dialects (16), and on the toneless penult in most OS dialects (17):

- (16) Final stress permitted
- |    |              |               |                 |
|----|--------------|---------------|-----------------|
| a. | /vɔ.dá/      | [vɔ.‘dá]      | ‘water.NOM.SG’  |
| b. | /pɔ.tók/     | [pɔ.‘tók]     | ‘creek.NOM.SG’  |
| c. | /vru.ɕei.ná/ | [vru.ɕei.‘ná] | ‘heat.NOM.SG’   |
| d. | /za.saa.dí/  | [za.saa.‘dí]  | ‘plant.IMP.2SG’ |
- (17) Final stress prohibited
- |    |              |              |                 |
|----|--------------|--------------|-----------------|
| a. | /vɔ.dá/      | [‘vɔ.dá]     | ‘water.NOM.SG’  |
| b. | /pɔ.tók/     | [‘pɔ.tók]    | ‘creek.NOM.SG’  |
| c. | /vru.ɕei.ná/ | [vru.ɕei.ná] | ‘heat.NOM.SG’   |
| d. | /za.saa.dí/  | [za.‘saa.dí] | ‘plant.IMP.2SG’ |

<sup>1</sup> Alternatively, Zec & Zsiga (2010)’s analysis of the NS word prosodic system is foot-free.

In OS dialects that prohibit final stress (17), only *light* final syllables are barred from bearing stress (18), while final heavies regularly receive stress if High-toned underlyingly (19).<sup>2</sup> Recall that a High must be attached to the first mora of a heavy syllable.

(18) Final High-toned lights

- |    |           |           |                 |
|----|-----------|-----------|-----------------|
| a. | /vɔ.dá/   | [vɔ.dá]   | ‘water.NOM.SG’  |
| b. | /pɔ.tók/  | [pɔ.tók]  | ‘creek.NOM.SG’  |
| c. | /ruu.ká/  | [ruu.ká]  | ‘arm.NOM.SG’    |
| d. | /naa.ród/ | [naa.ród] | ‘people.NOM.SG’ |

(19) Final High-toned heavies

- |    |            |             |                |
|----|------------|-------------|----------------|
| a. | /vɔ.déε/   | [vɔ.ˈdéε]   | ‘water.GEN.SG’ |
| b. | /vɔ.dóom/  | [vɔ.ˈdóom]  | ‘water.INS.SG’ |
| c. | /ruu.kéε/  | [ruu.ˈkéε]  | ‘arm.GEN.SG’   |
| d. | /ruu.kóom/ | [ruu.ˈkóom] | ‘arm.INS.SG’   |

Word-final High-toned lights regularly receive stress in enclisis, which results in productive stress alternations (20).

- |      |    |          |          |              |             |
|------|----|----------|----------|--------------|-------------|
| (20) | a. | [vɔ.dá]  | ‘water’  | [vɔ.ˈdá=jε]  | ‘water is’  |
|      | b. | [jε.zík] | ‘tongue’ | [jε.ˈzík=mi] | ‘my tongue’ |
|      | c. | [ruu.ká] | ‘arm’    | [ruu.ˈká=mi] | ‘my arm’    |

The illicitness of stressed final lights but not of stressed final heavies in OS falls out from the moraic version of NONFINALITY.<sup>3</sup> Mora-oriented NONFINALITY penalises word-level prominence on the rightmost mora of the prosodic word ( $\omega$ ) (Hyde, 2007). Stressed final heavies do not violate the moraic version of NONFINALITY because they form a bimoraic foot headed by the syllable’s first mora, which is the penultimate mora of a word. Thus, when stress is on a final heavy, the prominence peak does not coincide with the word’s rightmost mora.

(21) NONFINALITY( $\mu, \omega, Ft$ )

No mora which is final in the prosodic word heads a foot.

The second point of variation in OS concerns the metrical treatment of toneless heavy syllables. Individual OS dialects differ with respect to whether they prefer stressing High-toned lights over toneless heavies (22), or vice versa (23).

(22) High-toned light prevails

- |    |                |                 |                 |
|----|----------------|-----------------|-----------------|
| a. | /u.raa.dí.fε/  | [u.raa.ˈdí.fε]  | ‘do.AOR.3PL’    |
| b. | /naa.ró.di.ma/ | [naa.ˈró.di.ma] | ‘people.DAT.PL’ |

(23) Toneless heavy prevails

- |    |               |                |              |
|----|---------------|----------------|--------------|
| a. | /u.raa.dí.fε/ | [u.ˈraa.dí.fε] | ‘do.AOR.3PL’ |
|----|---------------|----------------|--------------|

<sup>2</sup>For the time being, I ignore the patterning of tone and weight in toneless penultimate lights. See §2.4.

<sup>3</sup>The present approach to nonfinality effects in OS differs slightly from Zec & Zsiga (2022). The differences between the two approaches are beyond the scope of this paper.



- b. /naa.ró.di.ma/      ['naa.ró.di.ma]      'people.DAT.PL'

The pattern in (23) arises in response to WEIGHT-TO-STRESS PRINCIPLE (WSP; Prince, 1990), defined in (24). For (23) to arise, WSP must dominate HEAD-H.

- (24)      WEIGHT-TO-STRESS PRINCIPLE  
             Assess a violation for every unstressed heavy syllable.

The combinations of the metrification strategies in (16)–(17) and (22)–(23) give rise to four basic prosodic types in OS. In Type 1 (East Montenegro; Stevanović, 1933), stress invariably falls on the High-toned syllable (25a). In Type 2 (Southwest, Central and East Serbia, parts of Montenegro and Kosovo; Jović, 1968; Simić, 1972), stress generally falls on the High-toned syllable, except when the High-toned syllable is light and final (25b). In Type 3 (Lepetane in the Kotor Bay, Montenegro; Tomanović, 1935) High-toned syllables are stressed unless the immediately preceding syllable is heavy, in which case stress falls on the toneless heavy (25c). Type 4 (Smederevo–Vršac dialect cluster; Ivić, 1958) avoids stressing final High-toned lights and prefers stressing toneless heavies over High-toned lights (25d). In all environments not listed in (25), stress invariably falls on the High-toned syllable in all four types of OS dialects.

- (25)      a.    Type 1  
                  /vɔ.dá/                      [vɔ.'dá]  
                  /u.raa.dí.fɛ/              [u.raa.'dí.fɛ]  
             b.    Type 2  
                  /vɔ.dá/                      ['vɔ.dá]  
                  /u.raa.dí.fɛ/              [u.raa.'dí.fɛ]  
             c.    Type 3  
                  /vɔ.dá/                      [vɔ.'dá]  
                  /u.raa.dí.fɛ/              [u.'raa.dí.fɛ]  
             d.    Type 4  
                  /vɔ.dá/                      ['vɔ.dá]  
                  /u.raa.dí.fɛ/              [u.'raa.dí.fɛ]

The prosodic variation observed across individual OS dialects can be attributed to the interplay of three constraints in OS: HEAD-H, NONFINALITY and WSP. In Type 1 dialects (25a), HEAD-H dominates both NONFINALITY and WSP. In Type 2 dialects (25b), HEAD-H takes precedence over WSP, but ranks below NONFINALITY. Type 3 (25c) is the mirror image of Type 2: HEAD-H outranks NONFINALITY, but ranks below WSP. In Type 4 (25d), HEAD-H is dominated by both NONFINALITY and WSP. The present study is primarily concerned with Type 2 and 4 OS dialects.

Putting the pieces together, OS stress is TONE-DRIVEN and QUANTITY-SENSITIVE. The sensitivity to tone is evident from the preference for stressing syllables with a lexical High. Quantity-sensitivity manifests itself in several interactions. In Type 3 and 4 OS dialects, toneless heavy syllables attract stress over High-toned lights. The ban on final prominence in OS likewise makes reference to weight. There is an asymmetry between final High-toned lights, which are unstressable in Type 2 and 4 OS dialects, and final High-toned heavies, which regularly receive stress in all OS dialects. All descriptive

facts presented in this section can be unified by assuming that OS forms the MORAIC TROCHEE, the head mora of which is preferably High-toned (26).

(26) OS moraic trochee

( $\acute{\mu}_s \mu_w$ )

where:

$\mu_s$  = strong, foot-heading mora;  $\mu_w$  = weak, nonhead mora

$\acute{\mu}$  = High-toned mora;  $\mu$  = toneless mora

This footing preference explains the weight-stress and tone-stress interaction, along with nonfinality effects in OS in a unified fashion. The moraic trochee analysis sheds light on another central property of OS word prosody: the ban on rising contours. In OS, Highs must be attached to the first mora of a heavy syllable. Assuming the metrical structure in (26) for OS, stressed heavy syllables form a bimoraic trochee in this dialect group. Accordingly, the second mora of a stressed heavy constitutes the nonhead position of a bimoraic trochee. Thus, in rising contours, a High would be linked to the nonhead mora. Cross-linguistically, there is a dispreference for High-toned foot nonheads. This is formally captured by the \*NONHEAD-H constraint (de Lacy, 2002):

(27) \*NONHEAD-H

Assess a violation for every instance of a High-toned mora in the nonhead position of a foot.

In OS, \*NONHEAD-H penalises stressed heavies with a rising contour tone because they feature a High on the nonhead mora of a bimoraic trochee (see §3.2.2 for further discussion).

### 2.1.3. Interim summary

OS and NS have virtually identical underlying representations. The key distinction between the two dialect groups is that OS preferably stresses the High-toned syllable, while NS stresses the immediately preceding syllable:

(28)	UR	OS	NS	Gloss
	/sɛ.kí.ra/	[sɛ.'kí.ra]	['sɛ.kí.ra]	'axe.NOM.SG'
	/u.pɔ.rɛɛ.diiɸ/	[u.pɔ.'rɛɛ.diɸ]	[u.'pɔ.rɛɛ.diiɸ]	'compare.PRS.2SG'
	/vɔ.dóɔm/	[vɔ.'dóɔm]	['vɔ.dóɔm]	'water.INS.SG'
	/daa.náa/	[daa.'náa]	['daa.náa]	'day.GEN.PL'

OS is a MORAIC TROCHEE dialect, while NS exhibits the SYLLABIC TROCHEE. The footing differences between the two dialect groups are outlined in (29).

(29)	UR	OS	NS	Gloss
	/sɛ.kí.ra/	[sɛ.('kí.ra)]	[('sɛ.kí).ra]	'axe.NOM.SG'
	/u.pɔ.rɛɛ.diiɸ/	[u.pɔ.('rɛɛ).diɸ]	[u.('pɔ.rɛɛ).diiɸ]	'compare.PRS.2SG'
	/vɔ.dóɔm/	[vɔ.('dóɔm)]	[('vɔ.dóɔm)]	'water.INS.SG'
	/daa.náa/	[daa.('náa)]	[('daa.náa)]	'day.GEN.PL'

OS stress pattern is QUANTITY-SENSITIVE, with feet containing at most two moras. NS stress pattern is QUANTITY-INSENSITIVE: feet comprise two syllables regardless of their weight.

## 2.2. The minimal foot in BCMS

BCMS tolerates both degenerate (i.e. monomoraic) feet and feet headed by a toneless mora. Monomoraic feet with a toneless head mora, the combination of these two individually tolerable marked structures, is categorically prohibited. This gives rise to a typologically rare gap in the foot inventory of BCMS, outlined in (30) (see also Zec, 1999: 237).

- (30) Moraic foot inventory of BCMS  
       Monomoraic feet:   ( $\acute{\mu}$ )     $*(\mu)$   
       Bimoraic feet:    ( $\acute{\mu}\mu$ )   ( $\mu\mu$ )

Effectively, bimoraic feet can incorporate two toneless moras or a High-toned mora followed by a toneless mora. However, the only mora of a degenerate foot must be High-toned. Across BCMS dialects, there is a widespread conspiracy against degenerate feet with a toneless head mora. Two strategies are employed to eliminate this illicit structure. First, BCMS displays a vowel lengthening process that targets toneless, but not High-toned, stressed lights (Zec, 1999) (§2.3). Second, there is a process of tonal flop in some OS dialects, which shifts Highs from unstressed syllables to underlyingly toneless stressed lights, but not to stressed toneless heavies (Ivić, 1958). Further evidence for the conspiracy against toneless degenerate feet is provided by the fact that many OS dialects that do not allow tonal flop regularly exhibit vowel lengthening in stressed toneless lights. Thus, I argue that tonal flop and vowel lengthening are competing strategies employed to eliminate toneless degenerate feet across different OS dialects of BCMS (§2.4).

## 2.3. Monosyllabic lengthening

Previous work on BCMS tone has identified two tonal classes in nominals and verbs: the High-toned class and the toneless class (Inkelas & Zec, 1988; Zec, 1999; Martinović, 2012). The two tonal classes observe different size restrictions: words that belong to the toneless class are minimally bimoraic, while High-toned words are allowed to be monomoraic. Different minimality conditions in the two classes fall out from a productive monosyllabic lengthening (ML) process which targets underlyingly toneless monomoraic content words (31), but not their High-toned counterparts (32) (cf. Zec, 1999).<sup>4</sup> The lengthening pattern in (31) is observed in virtually all Štokavian dialects with contrastive vowel quantity, including both NS and OS dialects (Kapović, 2015: 620).

- (31) ML in toneless monosyllables

<sup>4</sup> All toneless monosyllables with an underlyingly short vowel are treated as monomoraic and hence are subject to ML, which shows that final consonants do not contribute weight in BCMS.

- |    |          |                     |           |                     |
|----|----------|---------------------|-----------|---------------------|
| a. | [ˈlɛɛd]  | ‘ice.NOM.SG’        | [ˈlɛ.da]  | ‘ice.GEN.SG’        |
| b. | [ˈkɛɛɛr] | ‘daughter.ACC.SG’   | [ˈkɛɛ.ri] | ‘daughter.GEN.SG’   |
| c. | [ˈkɔɔst] | ‘bone.NOM.SG’       | [ˈkɔ.sti] | ‘bone.GEN.SG’       |
| d. | [ˈbɔɔs]  | ‘barefoot.NOM.SG.M’ | [ˈbɔ.si]  | ‘barefoot.NOM.PL.M’ |

## (32) No lengthening in High-toned monosyllables

- |    |          |                      |            |                      |
|----|----------|----------------------|------------|----------------------|
| a. | [ˈbrát]  | ‘brother.NOM.SG’     | [ˈbrá.ta]  | ‘brother.GEN.SG’     |
| b. | [ˈdéd]   | ‘grandfather.NOM.SG’ | [ˈdɛ.da]   | ‘grandfather.GEN.SG’ |
| c. | [ˈzdráv] | ‘healthy.NOM.SG.M’   | [ˈzdrá.ʋi] | ‘healthy.NOM.PL.M’   |
| d. | [ˈlíx]   | ‘pour.AOR.1SG’       | [ˈlí.smɔ]  | ‘pour.AOR.1PL’       |

The alternation in (31) results from vowel lengthening in monosyllabic forms rather than vowel shortening in the initial syllable of polysyllabic forms. The data in (33) indicate that there is no vowel shortening in the initial syllable of BCMS polysyllables, refuting the shortening analysis of (31).

## (33) Long stem-vowel: no shortening

- |    |          |               |            |               |
|----|----------|---------------|------------|---------------|
| a. | [ˈʋuuk]  | ‘wolf.NOM.SG’ | [ˈʋuu.ka]  | ‘wolf.GEN.SG’ |
| b. | [ˈgraað] | ‘town.NOM.SG’ | [ˈgraa.da] | ‘town.GEN.SG’ |
| c. | [ˈrɛɛf]  | ‘word.NOM.SG’ | [ˈrɛɛ.fɪ]  | ‘word.NOM.PL’ |

Cross-linguistically, monosyllabic lengthening processes are commonly driven by the requirement that the smallest permissible foot or prosodic word comprise at least two moras, stated as a violable constraint in (34).

## (34) FOOTBINARITY

All feet are binary on the moraic or syllabic level.

However, if BCMS ML applied in response to FtBIN alone, both (31) and (32) would be expected to undergo lengthening. Puzzlingly, ML makes crucial reference to tone: it is only toneless, and not High-toned monomoraic forms that undergo this process in BCMS.<sup>5</sup>

The dual behaviour of BCMS monosyllables illustrated in (31)–(32) can be analysed in two ways. On one account, vowel lengthening in monomoraic forms is the default. Lengthening is blocked in High-toned monosyllables because the resulting structure is marked. This is the gist of the analysis proposed by Zec (1999).

<sup>5</sup>It should be noted that the context of ML becomes non-surface-apparent as a result of a High tone insertion process. In BCMS, all underlyingly toneless prosodic words receive a High tone on their initial mora by a default insertion rule (recall (10)). The interaction between ML and Tone Insertion is illustrated in (i).

(i)	UR	/brát/ (32a)	/lɛd/ (31a)	/prɔpade/ (10a)
	Stress Assignment	ˈbrát	ˈlɛd	ˈprɔpade
	ML	N/A	ˈlɛɛd	N/A
	Tone Insertion	N/A	ˈlɛɛd	ˈprɔpade
	Surface Form	[ˈbrát]	[ˈlɛɛd]	[ˈprɔpade]

Tone Insertion counterbleeds ML: were Tone Insertion to apply first, toneless monomoraic forms would be rendered indistinguishable from High-toned monosyllables, which are exempt from lengthening. This interaction may be ascribed to contrast preservation (Lubowicz, 2003), whereby lengthening in toneless monosyllables overapplies to preserve the contrast between the two tonal classes in monomoraic stems. The overapplication of ML is beyond the scope of the present paper.

Another possibility, which this paper argues for, is that the absence of lengthening in High-toned monosyllables is the default. Toneless monomoraic feet constitute a uniquely marked structure, which triggers the otherwise inactive lengthening process.

#### 2.4. Tonal Flop and Penultimate Lengthening

In addition to ML, there are other processes that conspire against toneless monomoraic feet across BCMS dialects. Some OS dialects exhibit foot optimisation strategies that remove toneless monomoraic feet in polysyllables. These involve Tonal Flop (TF; Ivić, 1958; Simić, 1972; Remetić, 1985) and Penultimate Lengthening (PL; Ivić, 1958; Remetić, 1985). In this section, I provide a unified account of these two previously unaddressed metrical processes in OS.

In OS, stress generally falls on the word's only High-toned syllable. However, in Type 2 and Type 4 OS dialects (§2.1.2), stress falls on the toneless penult if the High-toned syllable is light and final. Curiously, in OS dialects that prohibit final prominence, toneless penultimate lights behave differently under stress from toneless penultimate heavies. This is reminiscent of the dual patterning of monosyllables seen in (31)–(32). Stressed toneless heavies consistently surface faithfully in all Type 2 and Type 4 OS dialects (Jović, 1968: 26; Simić, 1972: 51–54; Remetić, 1985: 40), as (35) shows. (Hereon feet are parenthesised.)

- |      |    |             |                   |                   |
|------|----|-------------|-------------------|-------------------|
| (35) | a. | /ruu.ká/    | [('ruu).ká]       | 'arm.NOM.SG'      |
|      | b. | /nii.sám/   | [('nii).sám]      | 'not.am'          |
|      | c. | /za.saa.dí/ | [za.('saa).dí]    | 'plant.IMP.2SG'   |
|      | d. | /pɔ.mɛɛ.ní/ | [pɔ.('mɛɛ).ní]    | 'mention.IMP.2SG' |
|      | e. | /u.zɛɛ.ǰé/  | [u.('zɛɛ).ǰé]     | 'take.AOR.3PL'    |
| (36) | a. | /ruu.ká/    | [ruu.('ká=mi)]    | 'arm=my'          |
|      | b. | /nii.sám/   | [nii.('sám=sɛ)]   | 'not.am=REFL'     |
|      | c. | /za.saa.dí/ | [za.saa.('dí=ga)] | 'plant=it'        |
|      | d. | /pɔ.mɛɛ.ní/ | [pɔ.mɛɛ.('ní=me)] | 'mention=me'      |
|      | e. | /u.zɛɛ.ǰé/  | [u.zɛɛ.('ǰé=ni)]  | 'took=us'         |

Unlike penultimate heavies, stressed penultimate lights undergo TF or PL in various OS dialects.

TF shifts High tone from unstressed final syllables to the preceding light syllable (37).<sup>6</sup> Enclisis reveals the original position of tone (38).

- |      |    |              |                 |                 |
|------|----|--------------|-----------------|-----------------|
| (37) | a. | /vɔ.dá/      | [('vɔ.da)]      | 'water.NOM.SG'  |
|      | b. | /dɔ.ǰlí/     | [('dɔ.ǰli)]     | 'came.M.PL'     |
|      | c. | /ɔ.táts/     | [('ɔ.tats)]     | 'father.NOM.SG' |
|      | d. | /grɛ.ɔ.tá/   | [grɛ.('ɔ.ta)]   | 'shame.NOM.SG'  |
|      | e. | /ɔ.ʒɛ.ní/    | [ɔ.('ʒɛ.ni)]    | 'marry.IMP.2SG' |
|      | f. | /sa.kri.vén/ | [sa.('krí.vɛn)] | 'hidden.M.SG'   |
| (38) | a. | /vɔ.dá/      | [vɔ.('dá=jɛ)]   | 'water=is'      |

<sup>6</sup>The TF data in (37) are from two sources: Simić, 1972: 47–48 and Remetić, 1985: 40–41.

b.	/dɔ.ʃlí/	[dɔ.(ʃlí=su)]	‘come=are.3PL’
c.	/ɔ.táts/	[ɔ.(ʔáts=mi)]	‘father=my’
d.	/gre.ɔ.tá/	[gre.ɔ.(ʔá=jɛ)]	‘it is shame’
e.	/ɔ.ʒɛ.ní/	[ɔ.ʒɛ.(ní=sɛ)]	‘get married’
f.	/sa.kri.vén/	[sa.kri.(ʔvén=jɛ)]	‘hidden=is’

The data in (37)–(38) exhibit a stress-dependent tonal alternation. In forms that display penultimate stress in response to NONFINALITY (37), High tone shifts to the stressed penult. However, the penultimate syllable remains toneless in enclisis, where stress falls on the originally High-toned syllable (38).

In a subset of OS dialects, PL is employed instead of TF. The counterparts of forms (37) exhibit no tone shift, as the lexical High tone remains on the light final syllable on the surface. Instead, the stressed toneless penult becomes heavy via PL (39).<sup>7</sup>

(39)	a.	/vɔ.dá/	[ʔ(vɔɔ).dá]	‘water.NOM.SG’
	b.	/dɔ.ʃlí/	[ʔ(dɔɔ).ʃlí]	‘came.M.PL’
	c.	/ɔ.táts/	[ʔ(ɔɔ).táts]	‘father.NOM.SG’
	d.	/gre.ɔ.tá/	[gre.ʔ(ɔɔ).tá]	‘shame.NOM.SG’
	e.	/ɔ.ʒɛ.ní/	[ɔ.ʔ(ʒɛɛ).ní]	‘marry.IMP.2SG’
	f.	/sa.kri.vén/	[sa.ʔ(krii).vén]	‘hidden.M.SG’

In enclisis (40), the light ultima receives stress, while the penult, which is unstressed in this context, displays a short vowel.

(40)	a.	/vɔ.dá/	[vɔ.(ʔdá=jɛ)]	‘water=is’
	b.	/dɔ.ʃlí/	[dɔ.(ʔlí=su)]	‘come=are.3PL’
	c.	/ɔ.táts/	[ɔ.(ʔáts=mi)]	‘father=my’
	d.	/gre.ɔ.tá/	[gre.ɔ.(ʔá=jɛ)]	‘it is shame’
	e.	/ɔ.ʒɛ.ní/	[ɔ.ʒɛ.(ʔní=sɛ)]	‘get married’
	f.	/sa.kri.vén/	[sa.kri.(ʔvén=jɛ)]	‘hidden=is’

This alternation in vowel quantity must be attributable to lengthening in the stressed position rather than shortening in the pre-tonic position. The latter option is inconsistent with the treatment of long vowels in pre-tonic syllables in OS. Underlyingly long vowels remain long in pre-tonic syllables in OS dialects, as indicated by forms like [ruu.(ʔká=mi)] ‘my arm’ in (38a).

Like ML, TF and PL do not apply across the board, but are rather restricted to stressed toneless lights. TF and PL therefore constitute additional evidence for the prohibition on toneless degenerate feet in BCMS.

<sup>7</sup>The PL data in (39) are from Remetić (1985): 40–44.

## 2.5. Conspiracy

TF and PL have received considerable attention in BCMS dialectology, primarily as classification criteria for OS dialects (Ivić, 1958; Remetić, 1985; Ligorio, 2016).<sup>8</sup> However, no explanatory account of these processes has been offered to date.

TF and PL can profitably be conceived of as foot-optimisation strategies. TF makes the foot-heading mora High-toned, bringing the foot in compliance with HEAD-H (15). Similarly, PL makes the foot bimoraic, as dictated by FtBIN (34). However, the purpose of TF and PL goes beyond the removal of a singly-marked structure in much the same way as ML.

TF shifts High tone from unstressed syllables to stressed toneless lights (37). By contrast, Highs do not shift to stressed toneless heavies, as can be seen in (35).

In the same vein, PL targets *toneless* stressed lights (39). However, no vowel lengthening is observed in High-toned stressed lights. Consider the data from Gallipoli Serbian (Ivić, 1957), a Type 2 OS dialect that exhibits PL, in (41).

- |      |    |             |              |                  |
|------|----|-------------|--------------|------------------|
| (41) | a. | /kó.maa.ta/ | '(kó).maa.ta | 'piece.GEN.SG'   |
|      | b. | /pó.jaa.sɛ/ | '(pó).jaa.sɛ | 'belt.ACC.PL'    |
|      | c. | /í.ɕaa.fɛ/  | '(í).ɕaa.fɛ  | 'go.IMPF.2/3SG'  |
|      | d. | /mó.laa.fɛ/ | '(mó).laa.fɛ | 'beg.IMPF.2/3SG' |

In (41), stress falls on a High-toned light syllable, which forms a monomoraic foot.<sup>9</sup> These stressed lights do not undergo vowel lengthening, unlike their counterparts in (39). The difference between (39) and (41) is that the lengthened stressed lights in (39) are toneless, while the stressed lights in (41), which are not subject to lengthening, are High-toned.

TF and PL thus bear a striking resemblance to ML: neither TF nor PL apply across the board, but are restricted to a specific, doubly-marked environment—stressed toneless lights. TF and PL do not eliminate toneless feet or degenerate feet in isolation, but specifically target toneless degenerate feet.

I propose that the purpose of TF and PL is to eliminate doubly-marked toneless degenerate feet, analogous to the ML process discussed in §2.3. Therefore, TF, PL and ML all take part in a CONSPIRACY (Kisseberth, 1970) against toneless degenerate feet in NS and OS. The present analysis draws a correlation between three superficially distinct metrical processes in BCMS, highlighting their functional unity.

ML always applies to toneless monomoraic forms in all OS dialects, as elsewhere in BCMS. In a subset of OS dialects, either TF or PL applies to toneless monomoraic feet in polysyllabic forms. TF and PL are in complementary distribution: they never cooccur within the same OS dialect. Therefore, ML eliminates toneless monomoraic feet in monosyllables, while either TF or PL eliminates this impermissible structure in polysyllables.

<sup>8</sup>The traditional terms for TF and PL in BCMS dialectology are METATAXIS (BCMS *metataksa*) and KANOVIAN LENGTHENING (BCMS *kanovačko duženje*), respectively; see Ligorio, 2016 for terminology and an accessible overview of the processes.

<sup>9</sup>On the following assumptions: (i) in OS, feet are maximally bimoraic, which prevents the peninitial heavy from being included in the foot, and (ii) OS respects syllable integrity, which ensures that the first mora of the peninitial syllable cannot be included in the foot to the exclusion of the syllable's second mora.

Crucially, OS and NS dialect groups share in the ban on toneless monomoraic feet despite the differences in the types of footing they employ (recall §2.1.3). OS is a quantity-sensitive moraic trochee dialect. By contrast, NS is a quantity-insensitive syllabic trochee dialect. NS consistently forms disyllabic feet in polysyllabic forms. Underlyingly monomoraic content words therefore constitute the only case in which monomoraic feet can arise and thus the only context in which *FtBIN* can be violated in NS. This explains why ML is the only process employed in NS to eliminate toneless degenerate feet: NS never forms degenerate feet outside monomoraic forms, consistent with the syllabic trochee analysis. Feet are maximally bimoraic in OS, resulting in cases where monomoraic feet can arise even in polysyllabic forms (see §3.2). Toneless degenerate feet occur in both monosyllables and polysyllables in OS, and thus all three processes argued here to militate against toneless degenerate feet (ML, TF and PL) are employed across OS dialects.

The significance of the OS data adduced in (37)–(39) is twofold. First, these data indicate that the prohibition against toneless degenerate feet in BCMS goes beyond the process of ML. Second, these data provide a straightforward answer to one of the outstanding questions regarding prosodic minimality in BCMS: is there a difference between the minimal prosodic word ( $\omega_{min}$ ) and the minimal foot ( $Ft_{min}$ ) of the language?

Following McCarthy & Prince (1986), it has been hypothesised that the minimal prosodic word is universally coextensive with the minimal foot. However, in many languages, there is a mismatch between the minimal prosodic word and the minimal foot (Downing, 1998; Garrett, 1999). Based solely on the ML data (31), it is impossible to determine whether ML arises due to foot minimality or a separate word minimality constraint. The OS data presented in (37)–(39) show unequivocally that the complex lengthening pattern in (31) is not restricted to monosyllables, and thus cannot instantiate an independent word minimality effect.

In BCMS, there is a significant relationship between tone and prosodic minimality requirements. Lexical tone not only interacts with stress, but also takes part in defining the minimal acceptable foot in the language. Although OS and NS exhibit different types of footing, they both share the conspiracy against toneless monomoraic feet.

### 3. Proposal

#### 3.1. Joint and independent activity of *HEAD-H* and *FtBIN* in BCMS

The upshot of §2 is that both monomoraic feet and feet with a toneless head mora are individually permitted, but toneless monomoraic feet are strictly prohibited in BCMS. Various dialects of BCMS employ a series of foot-optimisation strategies to eliminate toneless monomoraic feet. This prosodic minimality effect arises via a complex interplay between constraints on the permissible size of metrical constituents and those on tone-stress interaction.

Since BCMS is a trochaic language with tone-driven stress (Zec, 1999 a.o.), I customise Prince (1990)'s Trochaic Rhythmic Harmony Scale by incorporating into it tonal restrictions on foot-heading moras:



- (42) Foot Harmony Scale (BCMS)  
 $(\acute{\mu}_s\mu_w) \succ \{(\mu_s\mu_w), (\acute{\mu}_s)\} \succ *(\mu_s)$

Table 1 provides a summary of the foregoing discussion, listing different types of moraic trochees, their violation profiles and attestedness in BCMS.

**Table 1.** *Moraic trochees and their violation profiles (✓ = satisfies the constraint in question; ✗ = violates the constraint in question)..*

	HEAD-H	FtBIN	Permitted?	Example
$(\acute{\mu}_s\mu_w)$	✓	✓	yes	[ <sup>h</sup> (kráaʌ)] ‘king’ (72)
$(\mu_s\mu_w)$	✗	✓	yes	[ <sup>h</sup> (ruu).ká] ‘arm’ (35a)
$(\acute{\mu}_s)$	✓	✗	yes	[ <sup>h</sup> (brát)] ‘brother’ (32a)
$(\mu_s)$	✗	✗	no	N/A

Individual HEAD-H violations are tolerable across BCMS dialects, given numerous instances of feet headed by a toneless mora (39). Likewise, BCMS allows monomoraic feet with a High-toned head mora (32); (41). Accordingly, FtBIN is independently violable in the language.

Although the independent violations of HEAD-H and FtBIN are allowed in BCMS, coincident violation of these two markedness constraints is strictly prohibited. This explains the absence of toneless monomoraic feet from the language’s foot inventory (30). To avoid doubly-marked toneless monomoraic feet, BCMS subjects underlyingly toneless monomoraic feet to vowel lengthening (ML or PL) or TF. Both processes are otherwise hindered by faithfulness in BCMS.

I propose that the ban on doubly-marked toneless monomoraic feet in BCMS instantiates a GANG EFFECT (Jäger & Rosenbach, 2006; Pater, 2009b, 2016; Breiss, 2020; Breiss & Albright, 2022), whereby the combination of two individually tolerable marked structures is strictly disallowed. The purpose of ML, TF and PL is to remove the coincident violation of HEAD-H and FtBIN incurred by a toneless degenerate foot. ML, TF and PL do not aim to eliminate degenerate feet across the board, but target those degenerate feet which contain a toneless mora. This explains why these processes are restricted to stressed toneless light syllables and fail to apply in other marked environments. The relevant markedness constraints (HEAD-H and DEP- $\mu$ ) are individually overridden by faithfulness. However, in doubly-violating environments, these two constraints gang up to overcome a stronger faithfulness constraint.

The minimal foot in BCMS is shaped by the interaction of two markedness constraints: HEAD-H and FtBIN. Both constraints involved in this cumulative effect are independently active in BCMS phonology.

Recall from §2 that HEAD-H plays a pivotal role in OS, as High tone generally attracts stress in this dialect group. By contrast, in NS dialects, HEAD-H ranks relatively low because stress does not coincide with High tone, except in initial syllables (9)–(10). However, even in NS, HEAD-H is active in some corners of the grammar. Toneless input forms display initial stress and receive a High in the stressed syllable (10). This

indicates that although HEAD-H ranks too low in NS to consistently enforce tone-stress alignment, it lures inserted Highs towards foot-heading syllables by virtue of the emergence of the unmarked (McCarthy & Prince, 1994).

The independent effect of FTBIN in BCMS is somewhat difficult to evaluate outside the context of prosodic minimality. However, Werle (2009) demonstrates that FTBIN is active in two other areas of BCMS prosodic phonology: pitch accent assignment (cf. Werle, 2009: 74–77) and the prosodic patterning of prepositions (Werle, 2009: 211). In addition, stress-tone adjacency in NS provides indirect evidence for FTBIN. As argued in §2.1.1, stress-tone adjacency in NS is the result of the dialect’s preference for disyllabic trochees (Bethin, 1998), which are formed in response to FTBIN.<sup>10</sup>


In sum, the minimal foot requirement in BCMS arises through the interaction of two markedness constraints: HEAD-H and FTBIN. Both constraints are independently motivated in the language.

### 3.2. OT analysis


#### 3.2.1. Monosyllabic Lengthening via local constraint conjunction

BCMS tolerates feet with a toneless head mora (33); (39). In bimoraic feet, toneless head moras cannot acquire a High by tone insertion or TF. The faithful realisation of toneless foot heads in this context suggests that HEAD-H is individually overridden by the relevant faithfulness constraints, including DEP-H, which prohibits the insertion of a High, and NOFLOP-TONE (Alderete, 2001), which militates against TF. The rankings are provided in tableaux (43)–(44).

(43)


	/vuuk/	DEP-H	HEAD-H
a.	 '(vuuk)		*
b.	'(vúuk)	*!	

(44)

	/ruu.ká/	NOFLOP-H	NONFIN	HEAD-H
a.	ruu. '(ká)		*!	
b.	'(rúu).ka	*!		
c.	 '(ruu).ká			*

BCMS likewise permits monomoraic feet (32). This fact suggests that FTBIN is dominated by the anti-lengthening faithfulness constraint DEP- $\mu$ , as demonstrated in tableau (45).

(45)

	/brát/	DEP- $\mu$	FTBIN
a.	 '(brát)		*
b.	'(bráat)	*!	

<sup>10</sup>See Milenković (2023) for an OT account of tone-stress adjacency in NS.

When violated simultaneously, HEAD-H and FtBIN gang up and override a higher-ranking faithfulness constraint: DEP- $\mu$  in the case of ML and PL, and NoFLOP-H in the case of TF. Classical OT is incapable of deriving gang effects without nonstandard devices (Pater, 2009a,b, 2016). The failed OT derivation of ML is provided in (46).

(46) Failed OT analysis of ML

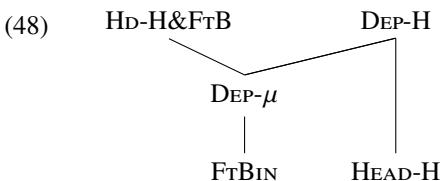
	/lɛd/	DEP-H	DEP- $\mu$	HEAD-H	FtBIN
a.	☛ <sup>*</sup> 'lɛd)			*	*
b.	'léd)	*!			*
c.	'léd)	*!	*		
d.	☺ 'lɛɛd)		*!	*	

OT accommodates ganging-up cumulativity using LOCAL CONSTRAINT CONJUNCTION (LCC; Smolensky, 1993, 2006), which incurs a violation whenever all conjuncts are violated within a designated local domain (e.g. syllable, foot, or the like). To capture the observed gang effect in BCMS, I introduce the local conjunction of HEAD-H and FtBIN at the foot level. The LCC analysis of vowel lengthening in toneless monosyllables is outlined in (47).

(47)

	/lɛd/	HD-H&FtBIN	DEP-H	DEP- $\mu$	HD-H	FtBIN
a.	'lɛd)	*!			*	*
b.	'léd)		*!			*
c.	'léd)		*!	*		
d.	☺ 'lɛɛd)			*	*	

The candidate that obeys DEP- $\mu$  wins out when FtBIN is violated on its own. This is the case in (45), where the faithful candidate ['brát] violates FtBIN but obeys HEAD-H. Faithful realisation is preferred because there is no simultaneous violation of HEAD-H and FtBIN. However, when HEAD-H and FtBIN are jointly violated, they gang up against DEP- $\mu$  (47). There are two ways to avoid the coincident violation of HEAD-H and FtBIN: vowel lengthening (candidate (d) in (47)) or High insertion (candidate (b) in (47)). Vowel lengthening is chosen over tone insertion because DEP-H outranks DEP- $\mu$  (cf. also Zec, 1999: 245, especially tableau (57)). The constraint grammar that derives ML is outlined in (48).



HEAD-H alone is indeterminate between the winner ['lɛɛd] and the faithful loser \*['lɛd], since both candidates violate HEAD-H. However, the conjoined constraint


HEAD-H&FTBIN favours unfaithful mapping over faithful realisation in (47). The winning candidate [ˈlɛɛd] fixes one (rather than both) markedness penalties incurred by its faithful contender. Nevertheless, the partial repair offered by the lengthened form is still sufficient to satisfy HEAD-H&FTBIN. This renders vowel lengthening optimal for toneless monomoraic inputs. Thus, the ML process is driven jointly by HEAD-H and FTBIN. In strict-ranking OT, this cumulative constraint interaction is captured via local constraint conjunction, as demonstrated in (47).

### 3.2.2. Tonal Flop and Penultimate Lengthening: constraint-based analysis

The analysis proposed in §3.2.1 readily extends to the processes of TF and PL observed in some OS dialects of BCMS. These are argued in §2.5 to have the same purpose as ML: the elimination of toneless degenerate feet.

TF targets stressed toneless *lights* (37), but not stressed toneless *heavies* (35). The absence of TF in cases like /ruu.ká/ → [ˈ(ruu).ká] ‘arm.NOM.SG’ in (35a) (not \*[ˈ(rúu).ka]) indicates that HEAD-H cannot be the sole driving force behind TF, else it would apply across the board. Judging from the absence of TF in (35), HEAD-H must be individually overridden by NoFLOP-TONE (49).



(49) NoFLOP ≫ HEAD-H

	/ruu.ká/	NONFIN	NoFLOP	HEAD-H
a.	 ˈ(ruu).ká			*
b.	ruu.ˈ(ká)	*!		
c.	ˈ(rúu).ka		*!	

The constraint grammar in (49) predicts that TF be inhibited in all environments in OS. This is problematic because stressed toneless lights attract tone from a final light syllable in some OS dialects, as (37) shows.

In tableau (50), I outline a failed attempt to derive the TF pattern in (37) under the constraint rankings from (49). This grammar erroneously favours the faithful candidate with penultimate stress over the intended winner that undergoes TF. Note that I ignore the issue of foot structure in the candidate forms in (50).

(50)

	/vɔ.dá/	NONFIN	NoFLOP	HEAD-H
a.	vɔ.ˈdǎ	*!		
b.	 ˈvɔ.dǎ			*
c.	 ˈvɔ.da		*!	

Consequently, the mappings in (49) and (50) appear to be inconsistent with each other. Example (49) points to NoFLOP ≫ HEAD-H. However, this ranking fails to derive the correct winner in (50).

Looping back to the form [ˈvɔ.dǎ] in (50), there are two structural analyses consistent with it. One possibility is outlined in (51a): the form displays a bimoraic trochee with

a toneless head mora and a High-toned nonhead mora. Alternatively, in (51b), the final High-toned light is unfooted and the penult forms a toneless degenerate foot.


(51) Structural analyses of ['vɔ.dá]

- a. [( 'vɔ.dá )]
- b. [ ' (vɔ).dá ]

The structure in (51a) satisfies FtBIN at the expense of \*NONHEAD-H, which requires that the nonhead mora of a foot not be High-toned. As argued in §2.1.2, High-toned foot nonheads are prohibited in OS. This prohibition manifests itself in the dialect's ban on rising contours, but also in the illicitness of form (51a). Accordingly, it is better for High-toned moras in OS to be left unfooted than to be included in the nonhead position of a foot.<sup>11</sup> This preference in OS is consistent with de Lacy (2002)'s theory of tone-stress interaction, where \*NONHEAD-H penalises High-toned foot nonheads and no constraint punishes unfooted High-toned moras or syllables.

In (51b), adherence to \*NONHEAD-H results in a toneless monomoraic foot. This structure is penalised by the conjoined constraint HEAD-H&FtBIN. Therefore, both faithful candidates that exhibit penultimate stress violate some top-ranking constraint in OS. The bimoraic foot in (51a) is ruled out by \*NONHEAD-H. The form in (51b) fatally violates HEAD-H&FtBIN. TF offers a way to satisfy both of these constraints. This process makes the stressed penult High-toned, satisfying HEAD-H, and renders the final light syllable toneless. This enables the inclusion of the final syllable in the nonhead position of a bimoraic trochee, which benefits FtBIN without offending \*NONHEAD-H. By so doing, the TF candidate [( 'vɔ.dá )] outperforms all faithful competitors (52). TF is preferred over vowel lengthening because DEP-μ outranks NoFlop-TONE.

(52) TF

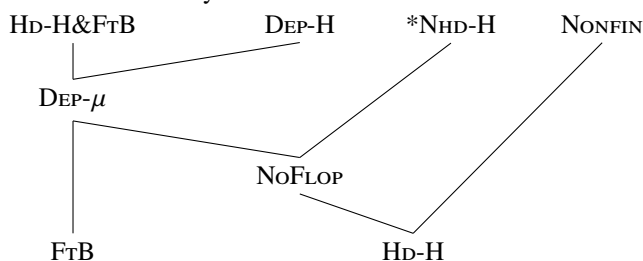
	/vɔ.dá/	H <sub>D</sub> -H&FtB	*N <sub>HD</sub> -H	N <sub>ONFIN</sub>	DEP-μ	NoFlop	H <sub>D</sub> -H	FtB
a.	vɔ.'(dá)			*!				*
b.	('vɔ.dá)		*!				*	
c.	'(vɔ).dá	*!					*	*
d.	'(vɔɔ).dá				*!		*	
e.	 ('vɔ.da)					*		

In sum, NoFlop cannot be violated in OS to comply with HEAD-H alone, as seen in (49). However, this constraint gives way to the local conjunction of HEAD-H and FtBIN (52).

OS dialects that exhibit TF likewise display vowel lengthening in toneless monomoraic words; cf. /ɛd/ → [lɛɛd] 'ice.NOM.SG'. The full constraint hierarchy for the subgroup of OS dialects that display TF is given in diagram (53).

<sup>11</sup>The behaviour of High-toned moras in OS parallels the pattern observed in Ayutla Mixtec (de Lacy, 2002: 11), where High-toned syllables are left unfooted to avoid High tone on the nonhead syllable of a disyllabic foot.

## (53) Constraint hierarchy for OS dialects with TF



The constraint hierarchy in (53) correctly derives both ML and TF. Since  $\text{DEP-}\mu$  dominates  $\text{NoFlop}$ , TF is preferred to vowel lengthening in polysyllabic forms (52). Vowel lengthening is the optimal strategy to eliminate toneless degenerate feet in monomoraic forms because  $\text{DEP-H}$  outranks  $\text{DEP-}\mu$ . Importantly, all crucial rankings relevant for ML (48) are included in (53).

The LCC analysis devised in (52) carries over to PL. This lengthening process makes the foot bimoraic, satisfying  $\text{FtBIN}$ . However,  $\text{FtBIN}$  cannot be the sole PL-inducing constraint. OS dialects tolerate monomoraic feet, as long as they are High-toned (54). This indicates that in OS,  $\text{FtBIN}$  ranks below  $\text{DEP-}\mu$ .

(54) Monomoraic feet tolerable in OS:  $\text{DEP-}\mu \gg \text{FtBIN}$ 

- a. /brát/ 'brother.NOM.SG' ['(brát)] not \*['(bráat)]  
 b. /kó.maa.ta/ 'piece.GEN.SG' ['(kó).maa.ta] not \*['(kóo).maa.ta]

Thus, PL targets toneless, but not High-toned, stressed lights. The process does not militate against all monomoraic feet, but rather *toneless* monomoraic feet.

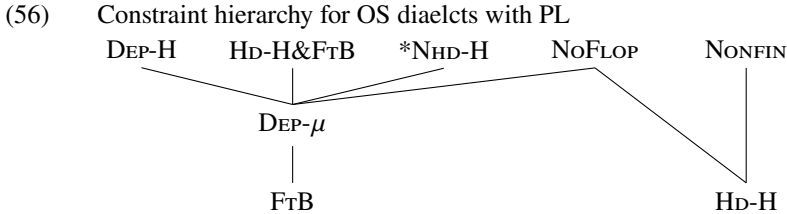
Importantly, all OS dialects display the metrical requirement that no foot be toneless and monomoraic simultaneously. All OS dialects employ ML in toneless monomoraic forms to avoid this doubly-marked structure. However, individual OS dialects employ different strategies to achieve this goal in disyllabic and polysyllabic forms. In some OS dialects, this is done by means of TF. These dialects rank  $\text{DEP-}\mu$  above  $\text{NoFlop}$ , as shown in (52). Conversely, PL arises in those OS dialects that rank  $\text{NoFlop}$  above  $\text{DEP-}\mu$ . The analysis is provided in (55).

## (55) PL

		HD-H&FtB	*NHD-H	NONFIN	NoFlop	DEP-μ	HD-H	FtB
a.	vɔ. '(dá)			*!				*
b.	('vɔ.dá)		*!				*	
c.	('vɔ).dá	*!					*	*
d.	('vɔ.da)				*!			
e.	ⱱⱱ ('vɔɔ).dá					*	*	

As in (52), final stress (candidate (a)) is ruled out in (55) by NONFINALITY. Candidate (b) is knocked out by \*NONHEAD-H. Candidate (c) fatally violates HEAD-H&FtBIN. The TF candidate (d) violates NoFLOP, which is undominated in this subgroup of OS dialects. The winning candidate (e) satisfies all top-ranking constraints at the expense of DEP- $\mu$ .

The sole difference between OS dialects that display PL and those that display TF lies in the ranking of DEP- $\mu$  relative to NoFLOP. The constraint hierarchy for the OS dialects that exhibit PL is given in (56).



In (56), NoFLOP  $\gg$  DEP- $\mu$  ensures that PL is preferred to TF in polysyllables. Similarly, DEP-H  $\gg$  DEP- $\mu$  ensures that ML is preferred to tone insertion in toneless monomoraic words. The constraint grammar in (56) retains all the constraint relations from (48) needed to derive ML.

Taking stock, I identified two subgroups of OS dialects. In one subgroup (52), there are two competing strategies to avoid toneless degenerate feet. In monosyllables, vowel lengthening is the optimal strategy to do away with a toneless degenerate foot. In polysyllables, TF takes precedence over vowel lengthening, ensured by the constraint hierarchy in (53). The other subgroup (55) consistently employs vowel lengthening to eliminate toneless degenerate feet, both in monosyllables and in polysyllabic forms. This preference falls out from the constraint hierarchy in (56).

I also provided a unified account of three superficially distinct metrical processes in BCMS: ML, TF and PL. These have not been considered functionally related in previous work. These processes take part in a conspiracy against toneless degenerate feet in the language. To capture the prohibition on toneless degenerate feet in BCMS, I introduced the local conjunction of HEAD-H and FtBIN.

### 3.3. HG analysis

This section focuses on how the gang effect observed in this paper is modelled in HG. Unlike OT, HG is capable of modelling gang effects without locally conjoined constraints (cf. Pater, 2009a,b, 2016; Farris-Trimble, 2008; Potts et al., 2010). However, the picture is more nuanced than this. Both OT and HG necessitate conjoined constraints to model superadditive ganging, whereby the effect of simultaneously violating two constraints exceeds the combination of effects of violating them individually (Albright, 2009; Green & Davis, 2014; Shih, 2017). This paper lends further support to this view. I show that the illicitness of doubly-marked toneless degenerate feet instantiates a superadditive gang effect. In line with this demonstration, the HG analysis of

prosodic minimality in BCMS does not dispense with the local conjunction of HEAD-H and F<sub>T</sub>BIN introduced earlier to model the effect of their coincident violation in OT.

### 3.3.1. Prosodic minimality in BCMS in standard HG

In the cases of cumulative markedness interaction typically discussed in the HG literature (Pater, 2009b, 2016), the repair strategy employed to remove coincident marked configurations eliminates *all* marked structures that are barred from cooccurring within a single domain. At the level of HG analysis, such cumulativity effects are modelled via ASYMMETRIC, i.e. ONE-TO-MANY, TRADE-OFFS between constraint weights (Pater, 2009b, 2016; Smith, 2022). Asymmetric trade-off between faithfulness (F) and markedness constraints M<sub>1</sub> and M<sub>2</sub> arises when the candidate that satisfies *either* M<sub>1</sub> *or* M<sub>2</sub> at the expense of F loses out to the faithful competitor, while the candidate that satisfies *both* M<sub>1</sub> *and* M<sub>2</sub> at the expense of F prevails against the faithful competitor.

This is how standard HG models the restrictions on voiced obstruents in Japanese loanword phonology, an oft-quoted example of ganging-up cumulativity (Nishimura, 2003; Kawahara, 2006; Pater, 2009b, 2016). In Japanese, multiple voiced obstruents are tolerated within a morpheme in loanwords, unlike in native vocabulary (57a) (Lyman, 1894). Likewise permitted in loanword phonology are otherwise intolerable voiced geminates (57b). However, when a voiced geminate cooccurs with a voiced singleton within a morpheme, the geminate optionally devoices (57c).


- (57)
- |    |         |         |
|----|---------|---------|
| a. | [gibu]  | ‘give’  |
|    | [bagi]  | ‘buggy’ |
| b. | [webbu] | ‘web’   |
|    | [heddu] | ‘head’  |
| c. | [gutto] | ‘good’  |
|    | [dokku] | ‘dog’   |

Since multiple voiced singletons are allowed within a morpheme (57a), IDENT-VOICE, which protects the underlying voicing of an obstruent, outweighs OCP-VOICE, which bars multiple voiced obstruents within a morpheme. Similarly, IDENT-VOICE outweighs NoVOICEDGEMINATE, which discriminates against voiced geminates, as evident from (57b). Geminate devoicing in (57c) is possible because OCP-VOICE and NoVOICEDGEMINATE gang up to overcome higher-weighted IDENT-VOICE.

Crucially, the repair employed to eliminate coincident marked structures in (57c) removes *both* relevant marked structures. This explains why it is the voiced geminate rather than the voiced singleton that undergoes devoicing. Singleton devoicing, as in \*[toggu], satisfies only OCP-VOICE, but not NoVOICEDGEMINATE. Thus, \*[toggu] shares a NoVOICEDGEMINATE violation with the faithful contender [doggu] and violates the individually strongest constraint (IDENT-VOI), as shown in (58).



## (58) Asymmetric trade-off between constraint weights



/doggu/		$\mathcal{H}$	Id-VOI	OCP-VOI	NoVCDGEM
			2	1.5	1.5
a.	doggu	-3		-1	-1
b.	 dokku	-2	-1		
c.	toggu	-3.5	-1		-1

Notably, in geminate devoicing in Japanese (58), a single unfaithful mapping satisfies *both* ganging constraints at the expense of a higher-weighted faithfulness constraint. This contrasts with the cumulativity effect in toneless monomoraic feet in BCMS, where no single repair strategy removes both marked structures. Both available repairs (vowel lengthening and tone insertion) remove only one of the coincident marked structures, while the other one persists.

Tone insertion (/ɛd/ → ['(léd)]) makes the foot-heading mora High-toned, thereby satisfying HEAD-H. However, the foot remains monomoraic, incurring a FTBIN penalty. Vowel lengthening (/ɛd/ → ['(lɛɛd)]) makes the foot bimoraic in response to FTBIN. The head mora remains toneless, violating HEAD-H. The only way to remove both of the coincident marked structures and satisfy both ganging constraints is to subject a toneless monomoraic foot to *both* vowel lengthening and tone insertion: ['(léɛd)]. This option violates *two* higher-weighted faithfulness constraints: DEP-H and DEP-μ.

Contrary to Japanese geminate devoicing (58), there is a SYMMETRIC, i.e. ONE-TO-ONE, TRADE-OFF between constraint weights in the HG analysis of ML in BCMS (59). For the intended winner ['(lɛɛd)] to prevail against the faithful contender ['(lɛd)], a higher-weighted DEP-μ penalty must be traded against a *single* lower-weighted penalty (FTBIN) in HEAD-H-violating contexts. An HG grammar with no means to amplify the severity of coincident constraint violations erroneously selects as optimal the faithful candidate for toneless monomoraic inputs. Due to the shared violation of HEAD-H, faithful realisation is favoured over lengthening regardless of the weight assigned to HEAD-H, as (59) shows.

## (59)

/ɛd/		$\mathcal{H}$	DEP-μ	HEAD-H	FTBIN
			2	$n$	1.5
a.	 ['(lɛd)]	$-(1.5 + n)$		-1	-1
b.	 ['(lɛɛd)]	$-(2 + n)$	-1	-1	

In (59), the intended winner (candidate (b)) satisfies only one of the markedness constraints, namely FTBIN, and still violates HEAD-H. Since the intended winner and the faithful contender (a) share a HEAD-H violation, vowel lengthening incurs a greater cumulative penalty than no lengthening. Having no means to add weight to the coincident violation of HEAD-H and FTBIN, the standard HG analysis cannot generate a gang effect under the weighting conditions in (59). These weighting conditions are independently motivated by the lack of lengthening in High-toned monosyllables in BCMS (32).

A similar result is observed in the PL process in OS. Under the weighting conditions independently established for OS, vowel lengthening in stressed toneless lights incurs a greater cumulative penalty than a doubly-marked toneless monomoraic foot. Because  $\text{DEP-}\mu$  is individually stronger than  $\text{FtBIN}$ , the weight incurred by coincident violation of  $\text{DEP-}\mu$  and  $\text{HEAD-H}$  is greater than that incurred by simultaneously violating  $\text{FtBIN}$  and  $\text{HEAD-H}$ . Under the standard HG analysis of PL (60), the intended winner (e) loses to the faithful candidate (c), which exhibits a toneless degenerate foot.

(60)

			NoFlop	DEP-H	NONFIN	*N <sub>HD</sub> -H	H <sub>D</sub> -H	DEP- $\mu$	FtBIN
/vɔ.dá/	$\mathcal{H}$		5	4	3	3	2	2	1.5
a.	vɔ. '(dá)	-4.5			-1				-1
b.	('vɔ.dá)	-5				-1	-1		
c.	☐ '(vɔ).dá	-3.5					-1		-1
d.	('vɔ.da)	-5	-1						
e.	☹ '(vɔvɔ).dá	-4					-1	-1	

Unlike the lengthening processes in (59) and (60), standard HG is capable of deriving TF. The HG analysis of TF is outlined in tableau (61).

(61)

			DEP- $\mu$	DEP-H	NONFIN	NoFlop	*N <sub>HD</sub> -H	H <sub>D</sub> -H	FtBIN
/vɔ.dá/	$\mathcal{H}$		5	4	3	3	2	2	1.5
a.	vɔ. '(dá)	-4.5			-1				-1
b.	('vɔ.dá)	-4					-1	-1	
c.	'(vɔ).dá	-3.5						-1	-1
d.	'(vɔvɔ).dá	-7	-1					-1	
e.	☐ ('vɔ.da)	-3				-1			

The weighting conditions in (61) are similar to those in (60), the main difference being the fact that  $\text{DEP-}\mu$  outweighs NoFlop in (61). As a result, the TF candidate (e) emerges as optimal in (61). TF satisfies both  $\text{HEAD-H}$  and  $\text{FtBIN}$  in one go, producing an asymmetric trade-off between NoFlop and two lower-weighted markedness constraints:  $\text{HEAD-H}$  and  $\text{FtBIN}$ . The HG analysis of TF accounts for this gang effect without any supplementary devices.

To recapitulate: standard HG is incapable of deriving tone-sensitive vowel lengthening in BCMS, wherein toneless monomoraic feet, but not their High-toned counterparts, become bimoraic to comply with the language's foot minimality condition. Under the

independently motivated weighting conditions in BCMS, the HG analysis of the relevant lengthening processes fails to replicate the effect of constraint conjunction in OT (47); (52); (55). This suggests that HG's built-in additive cumulativeness of constraint violations does not supplant all effects of conjoined constraints, which calls for the enrichment of standard HG's apparatus.

### 3.3.2. Weighted constraint conjunction in HG

The effect of violating  $\text{DEP-}\mu$  in BCMS is more severe than the effect of violating  $\text{FtBIN}$  individually. This is apparent from cases like (32), where High-toned monomoraic feet are unable to undergo vowel lengthening. The effect of coincident violation of  $\text{DEP-}\mu$  and  $\text{HEAD-H}$  is expected to exceed the effect of violating  $\text{FtBIN}$  and  $\text{HEAD-H}$  simultaneously, as the HG analysis in (59)–(60) encapsulates. The reasoning behind this expectation is schematised in (62).

- (62) a.  $e(\text{DEP-}\mu) > e(\text{FtBIN})$  (Observed)  
 b.  $e(\text{DEP-}\mu, \text{HEAD-H}) > e(\text{FtBIN}, \text{HEAD-H})$  (Expected)

However, what transpires in BCMS is inverse to (62b): the joint violation of  $\text{HEAD-H}$  and  $\text{FtBIN}$  has a more detrimental effect on metrical well-formedness than the coincident violation of  $\text{DEP-}\mu$  and  $\text{HEAD-H}$ , as (63) shows:

- (63)  $e(\text{FtBIN}, \text{HEAD-H}) > e(\text{DEP-}\mu, \text{HEAD-H})$  (Observed)

For (62a) and (63) to both be true, it must be the case that in doubly-violating contexts,  $\text{HEAD-H}$  and  $\text{FtBIN}$  add up to more than the combination of their independent contributions. In other words, the effect of the joint violation of  $\text{HEAD-H}$  and  $\text{FtBIN}$  is greater than the combination of the effects of these constraints' independent violations. Thus, the cumulative effect of  $\text{HEAD-H}$  and  $\text{FtBIN}$  violations in BCMS is SUPERADDITIVE rather than additive based on the criteria defined in (3):

- (64) Superadditivity of coincident  $\text{FtBIN}$  and  $\text{HEAD-H}$  violation  
 $e(\text{FtBIN}, \text{HEAD-H}) > e(\text{FtBIN}) + e(\text{HEAD-H})$

Following Green & Davis (2014) and Shih (2017), the HG analysis of prosodic minimality in BCMS must adopt the weighted local conjunction of  $\text{HEAD-H}$  and  $\text{FtBIN}$  to accommodate the exacerbated severity of their joint violation (64). The presence of this conjoined constraint in the HG grammar of BCMS adds weight to the coincident violation of  $\text{HEAD-H}$  and  $\text{FtBIN}$ . The extra weight attributed to the doubly-marked candidate (a) in tableau (65) enables the lengthening candidate (b) to prevail against the faithful competitor. The HG account of ML is provided in (65).<sup>12</sup>

<sup>12</sup>As the Associate Editor suggests, although both OT and HG need the local conjunction of  $\text{HEAD-H}$  and  $\text{FtBIN}$  to capture their interaction in BCMS, the way this conjoined constraint achieves superadditivity is different across the two frameworks. For superadditivity to arise in strict-ranking OT, the conjunction must outrank both conjuncts, as well as the constraint that the conjuncts gang against. This is the case in (47), where  $\text{HEAD-H} \& \text{FtBIN}$  outranks not only  $\text{HEAD-H}$  and  $\text{FtBIN}$ , but also  $\text{DEP-}\mu$ . In HG, superadditivity arises by the very presence of  $\text{HEAD-H} \& \text{FtBIN}$  in the grammar, as long as this conjoined constraint is assigned non-zero weight. This follows from the cumulativeness of constraint violations: the penalty incurred by the individual conjuncts counts towards Harmony, and the conjoined constraint makes this cumulative penalty more severe.

## (65) Monosyllabic Lengthening in HG

$/\text{ɛd}/$	$\mathcal{H}$	DEP- $\mu$	H <sub>D</sub> -H	H <sub>D</sub> -H&FtB	FtB
		2	$n$	2	1.5
a. $'(\text{ɛd})$	$-(3.5 + n)$		-1	-1	-1
b. $\text{ɛ}^{\text{35}}\text{ɛ}'(\text{ɛɛd})$	$-(2 + n)$	-1	-1		

The weighted LCC analysis in (65) readily extends to PL in OS dialects of BCMS, shown in (60) to pose a challenge to standard HG. With the added weight incurred by the joint violation of H<sub>HEAD</sub>-H and FtB<sub>IN</sub> (candidate (c) in (66)), subjecting the toneless light penult to vowel lengthening (e) emerges as the least costly metrical strategy under the weighting conditions in (66):

## (66)

$/\text{vɔ.dá}/$	$\mathcal{H}$	NoF <sub>LOP</sub>	DEP-H	NONFIN	*N <sub>H<sub>D</sub>-H</sub>	H <sub>D</sub> -H	H <sub>D</sub> -H&FtB	DEP- $\mu$	FtB <sub>IN</sub>
		5	4	3	3	2	2	2	1.5
a. $\text{vɔ.}'(\text{dá})$	-4.5			-1					-1
b. $'(\text{vɔ.dá})$	-5				-1	-1			
c. $'(\text{vɔ}).\text{dá}$	-5.5					-1	-1		-1
d. $'(\text{vɔ}.da)$	-5	-1							
e. $\text{ɛ}^{\text{35}}\text{ɛ}'(\text{vɔɔ}).\text{dá}$	-4					-1		-1	

While HG can derive TF in isolation without the need for LCC (61), this does not imply that OS dialects displaying TF do not require LCC for other relevant processes. Notably, all OS dialects with TF also exhibit ML in toneless monomoraic forms. Given that standard HG fails to account for ML (59), the local conjunction of H<sub>HEAD</sub>-H and FtB<sub>IN</sub> is necessary even in OS dialects with TF to capture the full range of foot minimality effects.

Therefore, the formal account of TF is fully compatible with that of ML: both processes can be captured under a single constraint grammar enriched with H<sub>HEAD</sub>-H&FtB<sub>IN</sub>, under the weighting conditions presented in (67). This is shown for TF in (67a). The same grammar generates ML in toneless monomoraic forms (67b), and no lengthening in their toneless counterparts (67c).

## (67) OS dialects with TF

a.

<i>/vɔ.dá/</i>	<i>ℋ</i>	DEP- $\mu$ 5	DEP-H 4	NONFIN 3	NoFLOP 3	*N <sub>HD</sub> -H 2	H <sub>D</sub> -H 2	H <sub>D</sub> -H&F <sub>T</sub> B 2	F <sub>T</sub> BIN 1.5
a. vɔ.'(dá)	-4.5			-1					-1
b. ('vɔ).dá)	-4					-1	-1		
c. '(vɔ).dá	-5.5						-1	-1	-1
d. '(vɔɔ).dá	-7	-1					-1		
e. $\mathbb{P}$ ('vɔ.da)	-3				-1				

b.

<i>/lɛd/</i>	<i>ℋ</i>	DEP- $\mu$ 5	DEP-H 4	NONFIN 3	NoFLOP 3	*N <sub>HD</sub> -H 2	H <sub>D</sub> -H 2	H <sub>D</sub> -H&F <sub>T</sub> B 2	F <sub>T</sub> BIN 1.5
a. '(lɛd)	-8.5			-1			-1	-1	-1
b. $\mathbb{P}$ ('lɛɛd)	-7	-1					-1		
c. '(léd)	-8.5		-1	-1					-1
d. '(léɛd)	-9	-1	-1						

c.

<i>/brát/</i>	<i>ℋ</i>	DEP- $\mu$ 5	DEP-H 4	NONFIN 3	NoFLOP 3	*N <sub>HD</sub> -H 2	H <sub>D</sub> -H 2	H <sub>D</sub> -H&F <sub>T</sub> B 2	F <sub>T</sub> BIN 1.5
a. $\mathbb{P}$ ('brát)	-4.5			-1					-1
b. '(bráat)	-5	-1							

In conclusion, the prosodic minimality effects in BCMS arise as a result of the interaction of HEAD-H and FTBIN. The cumulative effect of HEAD-H and FTBIN violations was shown to be superadditive. That being the case, the HG analysis of the processes that conspire against toneless degenerate feet necessitates the local conjunction of HEAD-H and FTBIN—just like the equivalent OT analysis advanced in §3.2.

OT and HG thus end up requiring the same machinery to model the interaction of HEAD-H and FTBIN in BCMS. This demonstration contributes to the OT–HG debate by providing further evidence for conjoined constraints in weighted constraint grammar

(Albright, 2009; Hayes et al., 2012; Green & Davis, 2014; Shih, 2017), contrary to the widespread assumption that HG obviates the need for conjoined constraints (Potts et al., 2010; Farris-Trimble, 2008; Pater, 2009a,b, 2016).

However, this study, considered in isolation, is indeterminate between OT and HG as a theory of ganging-up cumulativity, though other work has provided more conclusive evidence relevant to the OT-HG debate (see Zuraw & Hayes (2017); Breiss (2020); Smith & Pater (2020), a.o. for the current state of this debate). Importantly, the two frameworks invoke conjoined constraints for different purposes. OT avails itself of conjoined constraints to model all cases of ganging-up cumulativity, treating additive and superadditive ganging as equally complex in terms of the formal machinery they necessitate. By contrast, HG needs this supplementary device only for modelling superadditive ganging. HG thus treats superadditive ganging as more complex than additive ganging.

### 3.4. *No adverse effects of constraint conjunction*

The main contribution of HEAD-H&FtBIN is that it captures prosodic minimality in BCMS, which cannot be modelled with non-conjoined constraints only (§3.2–§3.3). This section investigates whether this conjoined constraint has any adverse effect on BCMS prosody beyond those corners of BCMS prosodic grammar that it was introduced to model. I algorithmically checked whether HEAD-H&FtBIN is consistent with the rest of the prosodic grammar of BCMS. The survey focused on OS, given that this dialect group displays more prosodic diversity than NS and employs all three processes discussed in this paper: ML, TF and PL.

The simulations were performed using the OT-Help software (Staubs et al., 2010), which uses the Recursive Constraint Demotion (RCD) algorithm (Tesar & Smolensky, 2000) to determine possible optima and calculate typology. To assess how much extra power HEAD-H&FtBIN adds to the phonological grammar of BCMS, I compared two constraint models: the Baseline model, which included eight independently motivated constraints (listed in Table 1 in the Appendix), and the LCC model, which included the eight Baseline constraints and the local conjunction of HEAD-H and FtBIN.<sup>13</sup>

The list of prosodic grammars generated by the OT-Help software was exported and subjected to further scrutiny using the R statistical programming environment (R Core Team, 2021). R script was created to aid in the comparison of the two constraint sets' predictions, and in the identification of input-output mappings inconsistent with the prosodic grammars of OS dialects.

The results showed that the Baseline model could not generate the ML pattern under any constraint ranking/weighting conditions. By contrast, the model enriched with HEAD-H&FtBIN generated 13 grammars that display this pattern. These grammars are listed in Table 3 in the Appendix.

Out of these 13 grammars, 8 matched known OS dialects. Notably, the remaining 5 grammars were not brought about by HEAD-H&FtBIN. For example, 4 of these 5 unattested grammars showed tonal flop to stressed heavies: /ruu.ká/ → ['(rúu).ka]. This pattern is not observed in OS dialects and can also be derived in the Baseline

<sup>13</sup>The input forms and candidates submitted to the OT-Help software are listed in Table 2 in the Appendix.

model with the ranking  $\text{HEAD-H} \gg \text{NoFLOP}$ . Similarly, the remaining unattested pattern (row 9 in Table 4) represents an accidental gap: this is a theoretically possible pattern that happens not to occur in the observed OS data. This pattern cannot be attributed to  $\text{HEAD-H\&FTBIN}$ .

In sum, the introduction of  $\text{HEAD-H\&FTBIN}$  achieves its intended purpose without introducing any effects that would not otherwise arise under the non-conjoined constraints independently motivated in BCMS prosody.

## 4. Alternatives

### 4.1. No uneven trochees: Zec 1999

On Zec (1999)'s account, vowel lengthening in toneless monosyllables in BCMS is driven by  $\text{FOOTBINARITY}$ , which is assumed to outrank  $\text{DEP-}\mu$ :<sup>14</sup>

(68)

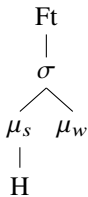
/lɛd/	FTBIN	DEP- $\mu$
a. 'lɛd)	*!	
b. 'lɛɛd)		*

To restrict ML to toneless monosyllables (31), Zec adopts Prince (1990)'s trochaic harmony scale, which asserts that symmetric trochees are preferred over uneven ones:

- (69) Trochaic harmony (Prince, 1990)  
 In a rhythmic unit (S W),  $|S| = |W|$   
 where  $|x|$  is the relative prosodic size of  $x$ .

Bimoraic feet formed by a single heavy syllable are left-headed (Prince, 1983; Kager, 1993).<sup>15</sup> Accordingly, the illicit lengthened form \*[bráat] 'brother' constitutes a moraic trochee, the structure of which is outlined in (70).

(70)



Per Zec (1999: 243–247) the structure in (70) is ruled out because it displays an uneven bimoraic foot, since its strong mora bears a High and its weak mora is toneless. The form thus fatally violates the size constraint dubbed  $\text{TROCHAICQUANTITY}$ , which requires equal metrical strength between the head and nonhead of a trochee.  $\text{TROCHQUANT}$  penalises  $(\acute{\mu}_s \mu_w)$  trochees for the same reason heavy-light trochees are considered

<sup>14</sup>Notably, ML is not the central concern of Zec (1999). The process, which is discussed in passing, is primarily intended to furnish additional evidence for the foot inventory proposed in the paper.

<sup>15</sup>Modulo marginal cases involving diphthongs with ascending sonority, which are considered to form moraic iambs (Kager, 1993).

disharmonic on Prince (1990)’s account. In BCMS, TROCHQUANT is satisfied at the expense of FTBIN, ensuring the faithful realisation of High-toned monomoraic inputs (71).

(71)

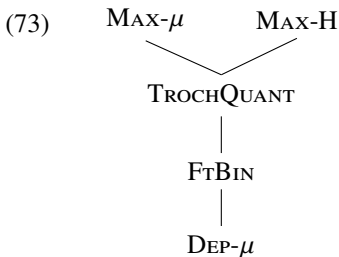
	/brát/	TROCHQUANT	FTBIN	DEP- $\mu$
a.	$\text{brát}$ ' (brát)		*	
b.	' (bráat)	*!		*

Moraic trochees with a High-toned head position are not categorically prohibited in BCMS, as can be seen in /kráaʌ/  $\rightarrow$  ['(kráaʌ)] ‘king’ in (72). This follows from the fact that TROCHQUANT ranks below MAX-H, which protects underlying Highs, and below MAX- $\mu$ , which militates against vowel shortening.

(72)

	/kráaʌ/	MAX-H	MAX- $\mu$	TROCHQUANT
a.	$\text{kráaʌ}$ ' (kráaʌ)			*
b.	' (kráaʌ)		*!	
c.	' (kraaʌ)	*!		

Thus, Zec (1999) captures the full range of relevant facts in BCMS monosyllables by the following constraint hierarchy:



To summarise: Zec (1999)’s account ascribes the absence of ML in High-toned monosyllables to the avoidance of the putatively marked ( $\acute{\mu}_s\mu_w$ ) trochees (71). Despite being disfavoured by TROCHQUANT, ( $\acute{\mu}_s\mu_w$ ) trochees are found elsewhere in the language because TROCHQUANT is dominated by faithfulness (72).

The chief difference between Zec (1999)’s account and the analysis proposed in this paper lies in what is considered the default state in BCMS monosyllables: vowel lengthening or the absence of lengthening. I argue that the absence of ML (32) is the default. Lengthening is generally inhibited except to eliminate toneless monomoraic feet. Therefore, the dual treatment of toneless and High-toned monosyllables in BCMS instantiates a TRIGGERING, i.e. DO-SOMETHING-ONLY-WHEN, effect (in the sense of Prince & Smolensky, 2004: §3–4). Conversely, Zec (1999) holds that vowel lengthening in BCMS monosyllables (31) is the default. It is only when ML results in a presumed illicit ( $\acute{\mu}_s\mu_w$ ) trochee that the process is blocked. This instantiates a BLOCKING, i.e. DO-SOMETHING-EXCEPT-WHEN, effect.



## 4.2. Theoretical and empirical issues with *TrochaicQuantity*

[Zec \(1999\)](#)'s account of the lengthening process in BCMS monosyllables remains uncontested to date in the South Slavic and prosodic literature. The account generates the lengthening pattern in BCMS monosyllables, but at a cost. In what follows, I discuss the theoretical implications and predictions of [Zec \(1999\)](#)'s account of ML, ultimately rejecting it. §4.2.1 takes up some questions surrounding the theoretical underpinning of this account. §4.2.2 and §4.2.3 show that it suffers from both overgeneration and undergeneration.

### 4.2.1. *TrochaicQuantity* and pitch-based rhythmic grouping

Central to [Zec \(1999\)](#)'s explanation of why ML fails to apply in High-toned monosyllables in BCMS is the preference for equal grouping in trochaic systems, ensuring that the foot-heading position is not heavier than the weak position ([McCarthy & Prince, 1986](#); [Prince, 1990](#); [Hayes, 1995](#)). [Zec \(1999\)](#) extends this requirement to pitch-based grouping. The *TROCHQUANT* constraint penalises the strong-weak contrast based on relative pitch, ensuring that the strong syllable/mora of a trochee must not be higher-pitched than the weak syllable/mora.

It has been observed that rhythmic groupings with varying intensity are preferably perceived as left-headed, while groupings with a contrast in duration are perceived as right-headed. The rationale behind the *TROCHQUANT* constraint is that grouping preferences based on pitch parallel those based on duration. There has been plentiful experimental work on rhythmic grouping effects (see [Crowhurst, 2020](#) for an overview). Many experimental studies on pitch-based grouping effects report a trochaic bias in alternating sequences with varying pitch ([Bion et al., 2011](#); [de la Mora et al., 2013](#); [Bhatara et al., 2013](#)). These results suggest that pitch-based grouping effects differ from duration-based grouping effects, given that higher-pitched elements are preferably associated with group onsets.

Therefore, pitch patterns with intensity rather than duration with respect to rhythmic grouping. The reported trochaic bias in pitch-based grouping poses a challenge for [Zec \(1999\)](#)'s account of BCMS ML, which holds that trochees with a higher-pitched first element are disfavoured in much the same way as those with a contrast in duration. Thus, one of the central theoretical underpinnings of [Zec \(1999\)](#)'s account—the parallelism between quantity-based and tone-based grouping preferences—is contradicted by available experimental evidence. This challenges [Zec \(1999\)](#)'s assertion that pitch behaves like duration in rhythmic grouping.

### 4.2.2. *TrochaicQuantity* penalises doubly-unmarked structure

*TROCHQUANT* treats bimoraic trochees headed by a High-toned mora as marked. This contradicts two prevalent cross-linguistic tendencies, one being the attraction of High tone to prosodically prominent positions ([Goldsmith, 1987](#); [Hayes, 1995](#); [Yip, 2001](#); [de Lacy, 2002](#); [Gordon, 2023](#)), and the other being the affinity between stress and heavy syllables ([Prince, 1990](#); [Hayes, 1995](#); [Gordon, 2006](#); [Ryan, 2016](#)). This constraint implies that the combination of two unmarked structures results in a marked configuration: both High-toned foot heads and bimoraic trochees are unmarked individually, but their combination incurs a markedness penalty.

Resultingly, TROCHQUANT produces a number of pathological effects. First, consider the constraint grammar in which TROCHQUANT outranks MAX- $\mu$ . In this hypothetical language, heavy syllables with a high-toned first mora become light when stressed (74). However, their toneless counterparts surface faithfully (75). In other words, a bimoraic trochee becomes degenerate because its head mora is High-toned.

(74)

	/ $\acute{\mu}\mu$ /	TRQ	MAX-H	MAX- $\mu$	FtBIN
a.	'( $\acute{\mu}\mu$ )	*!			
b.	'( $\mu\mu$ )		*!		
c.	'( $\acute{\mu}$ )			*	*

(75)

	/ $\mu\mu$ /	TRQ	MAX- $\mu$	FtBIN
a.	'( $\mu\mu$ )			
b.	'( $\mu$ )		*!	*!

Second, TROCHQUANT's dominance over MAX-H gives rise to a pathological tone deletion pattern whereby High tone deletes in the head position of a bimoraic trochee (76), but not in the head position of a monomoraic foot (77) and in unstressed syllables (78):

(76)

	/ $\acute{\mu}\mu$ /	TRQ	MAX- $\mu$	MAX-H
a.	'( $\acute{\mu}\mu$ )	*!		
b.	'( $\acute{\mu}$ )		*!	
c.	'( $\mu\mu$ )			*

(77)

	/ $\acute{\mu}$ /	TRQ	MAX-H
a.	'( $\acute{\mu}$ )		
b.	'( $\mu$ )		*!

(78)

	/ $\acute{\mu}.\mu\mu$ /	TRQ	FtB	MAX-H
a.	' $\acute{\mu}.$ '( $\mu\mu$ )			
b.	' $\mu.$ '( $\mu\mu$ )			*!
c.	'( $\acute{\mu}$ ). $\mu\mu$		*!	

Another adverse effect of TROCHQUANT is illustrated in (79). A High is lured away from the foot-heading mora to an unfooted mora to comply with TROCHQUANT. The pattern is counterintuitive because documented cases of tone shift observe the opposite tendency: Highs shift from metrically weak positions to prominent positions (Downing, 1990; Bickmore, 1995; de Lacy, 2002; Breteler, 2018).

(79)

	<i>/μ.μ̇μ/</i>	TROCHQUANT	MAX-H	NoFlop
a.	$\mu. '(\mu\mu)$	*!		
b.	$\mu. '(\mu\mu)$		*!	
c.	$\mu. '(\mu\mu)$			*

Finally, even when TROCHQUANT ranks below faithfulness and thus fails to induce any unfaithful mappings, it is still capable of producing pathological effects. Suppose there is a language with lexical tone that preferably stresses the leftmost heavy syllable, else defaulting to the leftmost syllable, as in (80). Suppose further that this hypothetical language ranks TROCHQUANT over WSP. This ranking gives rise to the pathological mapping in (80c): the leftmost heavy, the first mora of which is High-toned, repels stress to the light peninitial syllable for no reason other than having a High-toned first mora.

- (80)
- a. Stress leftmost heavy:  
 $\mu.\mu\mu.\mu \rightarrow [\mu. ' \mu\mu.\mu]$
  - b. Else leftmost syllable:  
 $\mu.\mu.\mu \rightarrow [ ' \mu.\mu.\mu]$
  - c. Do not stress leftmost heavy if High-toned:  
 $\mu\mu.\mu.\mu \rightarrow [\mu\mu. ( ' \mu.\mu)]$

This repulsion effect runs afoul of the propensity of High tone to attract stress (Goldsmith, 1987; Gordon, 2023).

#### 4.2.3. Undergeneration

Zec (1999) discusses ML without addressing the other processes that conspire against toneless degenerate feet: TF and PL. This section explores whether Zec (1999)'s account extends to these other functionally related processes in BCMS.

To accomplish this, I conducted a learning simulation using the OT-Help software. The constraint set utilised in this simulation mirrors the one used in §3.4, the only difference being the inclusion of Zec (1999)'s TROCHQUANT in place of the HEAD-H&FTBIN conjunction from §3.4.

The simulation found that Zec (1999)'s account captures only a subset of the relevant processes in BCMS. Specifically, TROCHQUANT struggles to model a subgroup of the OS dialects that display TF (37). The grammar that Zec (1999)'s constraint set fails to derive is given in (81). These data match the stress pattern of the Kosovo-Resava dialect of BCMS (Ivić, 1958; Simić, 1972).

- (81) Phonological grammar problematic for TROCHQUANT

a.	<i>/bɔg/</i>	→	[ ' (bɔg)]	'god.NOM.SG'
b.	<i>/brát/</i>	→	[ ' (brát)]	'brother.NOM.SG'
c.	<i>/vɔ.dá/</i>	→	[ ' (vɔ).da]	'water.NOM.SG'
d.	<i>/ruu.ká/</i>	→	[ ' (ruu).ká]	'arm.NOM.SG'
e.	<i>/u.raa.dí.ʃɛ/</i>	→	[u.raa. ' (dí).ʃɛ]	'do.AOR.3PL'

f.	/sɛ.kí.ra/	→	[sɛ.ʼ(kí).ra]	‘axe.NOM.SG’
g.	/prɔ.dáav.ni.tsa/	→	[prɔ.ʼ(dáav).ni.tsa]	‘store.NOM.SG’

The attentive reader will notice that the surface forms in (81c), (81e) and (81f) display monomoraic ([ʼ(vɔ́).da]) rather than bimoraic ([ʼ(vɔ́.da)]) feet. This adjustment was necessary in order for Zec (1999)’s TROCHQUANT-based account to be made to work, given that TROCHQUANT penalises bimoraic trochees with a High-toned head mora. On the present account, monomoraic feet posited in (81c), (81e) and (81f) would violate FTBIN for no obvious reason since HEAD-H&FTBIN does not favour monomoraic over bimoraic feet. However, for Zec (1999)’s account, it is crucial to assume that these OS forms exhibit monomoraic rather than bimoraic feet.

Even when this adjustment in favour of TROCHQUANT is made, the algorithm finds no OT grammar that derives all mappings in (81). The RCD algorithm stalled without reaching the target grammar. Six out of nine constraints were left unranked. The following constraints were left unranked after the RCD algorithm stalled: HEAD-H, FTBIN- $\mu$ , WSP, NoFLOP-H, DEP- $\mu$  and TROCHQUANT.

No totally ranked hierarchy was established for the above six constraints due to a ranking inconsistency involving TROCHQUANT. The RCD algorithm demotes DEP- $\mu$  below FTBIN on the basis of vowel lengthening in toneless monosyllables (81a). Since there is no lengthening in High-toned monomoraic forms (81b), the algorithm demotes FTBIN below TROCHQUANT. The subranking TROCHQUANT  $\gg$  FTBIN  $\gg$  DEP- $\mu$  is precisely how Zec (1999: 244–245) accounts for vowel lengthening in toneless monosyllables in BCMS, as discussed in §4.1.

The TF mapping in (81c) provides rationale to demote NoFLOP below HEAD-H. This is illustrated in the combination tableau in (82). The only way to justify the absence of TF in (81d) is by demoting HEAD-H below TROCHQUANT (83). Finally, WSP gets demoted below HEAD-H to accommodate the preference for stressed High-toned lights over stressed toneless heavies in the Kosovo-Resava dialect (81e), illustrated in tableau (84).

(82) Demote NoFLOP below HEAD-H

	/vɔ́.dá/	HEAD-H	NoFLOP
a.	☞ ʼ(vɔ́).da		1
b.	ʼ(vɔ́).dá	W1	L

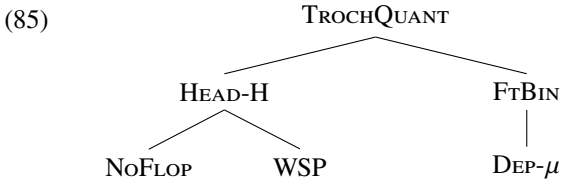
(83) Demote HEAD-H below TROCHQUANT

	/ruu.ká/	TROCHQUANT	HEAD-H
a.	☞ ʼ(ruu).ká		1
b.	ʼ(rúu).ka	W1	L

(84) Demote WSP below HEAD-H

	/u.raa.dí.fɛ/	HEAD-H	WSP
a.	☞ u.raa.ʼ(dí).fɛ		1
b.	u.ʼ(raa).dí.fɛ	W1	L

After these five iterations of constraint demotion, the learner posits the following constraint grammar:



The grammar in (85) is inconsistent with the mapping /prɔ.dáau.ni.tsa/ → [prɔ.ʼ(dáau).ni.tsa] (81g). Taken in isolation, (81g) requires HEAD-H to dominate TROCHQUANT. However, this ranking contradicts the ranking established in (83):

(86) RCD crashes

prɔ.dáau.ni.tsa			TRQU	HD-H	FTB	NoFlop	WSP	DEP-μ
a.	☺	prɔ.ʼ(dáau).ni.tsa	1					
b.	☹ <sup>a</sup>	prɔ.dáau.(ʼni.tsa)		1			1	

Thus, the RCD algorithm cannot find a totally ranked hierarchy for the six constraints in (85). Employing TROCHQUANT to model prosodic minimality in the OS dialects that display TF creates a ranking paradox. The mapping in (81g)/(86) cannot be made consistent with the rest of the dialect's phonological grammar in (81).

This undesirable effect of Zec (1999)'s TROCHQUANT is limited to strict-ranking OT. The inconsistency encountered in (86) does not carry over to HG, where HEAD-H and WSP can gang up to jointly override TROCHQUANT. This gang effect enables the intended winner [prɔ.ʼ(dáau).ni.tsa] to prevail against the problematic competitor \*[prɔ.dáau.(ʼni.tsa)] in (86).

### 4.3. Section summary

I have pointed out that Zec (1999)'s account of ML suffers from both overgeneration and undergeneration. The account overgenerates because the TROCHQUANT constraint predicts a number of pathological effects (§4.2.2). Zec (1999)'s constraint model likewise fails to generate some of the relevant processes in OS, at least within the bounds of strict-ranking OT (§4.2.3). Further, Zec (1999)'s account establishes no link between TF, PL and ML, which are evidently functionally related. Moreover, TROCHQUANT requires unmotivated auxiliary hypotheses, including the preference for monomoraic over bimoraic feet.

None of these adverse effects carry over to the LCC account proposed in this paper: it derives all relevant patterns in BCMS (§3.2–§3.3) and has no additional detrimental effects (§3.4). The ban on toneless degenerate feet falls out from the interaction of two constraints that are independently motivated in BCMS phonology: HEAD-H and FTBIN. Consequently, the present account obviates the need for ad hoc constraints à la Zec

(1999)'s TROCHQUANT to block vowel lengthening in High-toned monomoraic feet or inhibit TF in bimoraic feet. This effectively dispenses with the problematic theoretical assumptions and predictions of Zec (1999)'s account identified in this section.

## 5. Conclusion

This paper explored prosodic minimality in BCMS. A key finding was the identification of a relationship between tone and foot size in the language. Individually, BCMS permits both monomoraic feet and feet with a toneless head. However, the language disallows toneless monomoraic feet, the combination of these marked structures.

The present study contributes to our understanding of this complex prosodic pattern in two ways. First, I introduced new data from dialectal BCMS to a wider generative audience. These data show that vowel lengthening in toneless monosyllables, discussed in passing by Zec (1999), is not the sole manifestation of the ban on toneless monomoraic feet. In fact, there is a cross-dialectal conspiracy against this doubly-marked structure. Second, this study offers an alternative to the only existing generative account of prosodic minimality in BCMS (Zec, 1999), which makes flawed typological predictions and fails to account for the full range of minimality effects in BCMS in a unified fashion.

I attributed the ban on toneless degenerate feet to the joint effect of two markedness constraints independently motivated in the phonology of BCMS. These constraints include HEAD-H, which requires that foot-heading moras be High-toned, and FTBIN, which penalises monomoraic feet. I further demonstrated that the effect of the coincident violation of these two constraints is superadditive. I introduced the local conjunction of these constraints to capture the exacerbated severity of their joint violation in BCMS. The proposed conjoined constraint is necessary in both strict-ranking OT and, importantly, HG, which is often argued to fully dispense with conjoined constraints (Pater, 2009b; Farris-Trimble, 2008; Potts et al., 2010).

The main takeaways of the study are the following:

1. This study identifies a case of cumulative markedness interaction in BCMS whereby the combination of two individually tolerable marked structures is categorically ruled out. This pattern was shown to be superlinear: the effect of simultaneous violation of the relevant markedness constraints goes beyond the combination of their independent effects. This demonstration provides support for the use of locally conjoined constraints as a means of modelling superlinear ganging-up cumulativity in weighted constraint grammar (Albright, 2009; Hayes et al., 2012; Green & Davis, 2014; Shih, 2017).
2. Virtually all superadditivity effects have been documented in variable phonological patterns (Shih, 2017; Smith & Pater, 2020; Breiss & Albright, 2022; Kim, 2022). This study identifies a superadditive cumulativity effect in a categorical prosodic pattern, thereby expanding the empirical range of attested superadditivity effects.
3. The identification of a tone-sensitive prosodic minimality effect has implications for prosodic typology. In BCMS, lexical Highs not only determine the locus of stress, which is a well-known tendency in languages with tone-driven stress (de

Lacy, 2002; Gordon, 2023), but also take part in defining the minimal size of prosodic constituents.

4. The analysis has the added benefit of unifying three superficially distinct processes in BCMS—Monosyllabic Lengthening, Tonal Flop and Penultimate Lengthening—which had not yet been considered related.

**Acknowledgments.** I would like to thank an anonymous *Phonology* reviewer, reviewer Canaan Breiss and the Associate Editor, all of whose extensive comments led to substantial improvements of the initial submission. The paper greatly benefited from feedback by audience members at the Harvard PhonLab meeting, CLS 59, FASL 32, and AMP 2023, especially Michael Becker, Canaan Breiss, Aleksei Nazarov and Paul Smolensky. Special thanks go to Kevin Ryan for continuous support and a careful reading of the manuscript, and to Anabelle Caso for an outstanding job editing the revised manuscript. All remaining errors are solely my responsibility.

**Data availability statement.** Replication data and code for this study can be found in Harvard Dataverse: <https://doi.org/10.7910/DVN/1V64L2>.

**Supplementary material.** The appendix, which contains the three tables referenced in §3.4, is included in the supplementary material available at (link).

**Competing interests.** The author declares that there are no conflicts of interest regarding the publication of this paper.

**Ethical standards.** The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

## References

- Albright Adam (2009). *Cumulative violations and complexity thresholds: evidence from Lakota*. Paper presented at the 17th Manchester Phonology Meeting.
- Albright, Adam (2012). *Additive markedness interactions in phonology*. Unpublished manuscript, MIT.
- Alderete, John (2001). Dominance effects as transderivational anti-faithfulness. *Phonology* 18(2), 201–253.
- Baković, Eric (2000). *Harmony, dominance and control*. PhD dissertation, Rutgers University.
- Batas, Ana (2014). *Fonetska i akcenatska promenljivost reči u kontinualnom govoru*. PhD dissertation, University of Belgrade.
- Bethin, Christina (1994). On the phonology of the Neoštokavian accent retraction in Serbian and Croatian. *Die Welt der Slaven*, 39(2), 277–296.
- Bethin, Christina (1998). *Slavic prosody: Language change and phonological theory*. Cambridge: Cambridge University Press.
- Bhatara, Anjali, Natalie Boll-Avetisyan, Annika Unger, Thierry Nazzi & Barbara Höhle (2013). Native language affects rhythmic grouping of speech. *JASA* 134(5), 3828–3843.
- Bickmore, Lee (1995). Tone and stress in Lamba. *Phonology* 12(3), 307–341.
- Bion, Ricardo, Silvia Benavides-Varela & Marina Nespor (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and speech* 54(1), 123–140.
- Breiss, Canaan (2020). Constraint cumulativity in phonotactics: Evidence from artificial grammar learning studies. *Phonology* 37(4), 551–576.
- Breiss, Canaan & Adam Albright (2022). Cumulative markedness effects and (non-)linearity in phonotactics. *Glossa: a journal of general linguistics* 7(1).
- Breteler, Jeroen (2018). *A foot-based typology of tonal reassociation: Perspectives from synchrony and learnability*. PhD dissertation, University of Amsterdam.
- Crowhurst, Megan (2020). The iambic/trochaic law: Nature or nurture?. *Language and Linguistics Compass* 14(1).
- de Lacy, Paul (2002). The interaction of tone and stress in Optimality Theory. *Phonology* 19(1), 1–32.
- de la Mora, Daniela, Marina Nespor & Juan Toro (2013). Do humans and nonhuman animals share the grouping principles of the iambic–trochaic law?. *Attention, Perception, & Psychophysics*. 75(1), 92–100.

- Downing, Laura (1990). Local and metrical tone shift in Nguni. *Studies in African Linguistics* 21(3), 261–318.
- Downing, Laura (1998). On the prosodic misalignment of onsetless syllables. *NLLT* 16(1), 1–52.
- Farris-Trimble, Ashley (2008). *Cumulative faithfulness effects in phonology*. PhD dissertation, Indiana University.
- Garrett, Edward (1999). Minimal words aren't minimal feet. In Matthew Gordon (ed.) *Papers in Phonology 2: UCLA Working Papers in Linguistics* 1, 68–105.
- Goldsmith, John (1987). Tone and accent, and getting the two together. *BLS* 13, 88–104.
- Gordon, Matthew (2006). *Syllable weight: phonetics, phonology, typology*. New York: Routledge.
- Gordon, Matthew (2023). The Phonetic Basis for Tone-Stress Interactions: A Cross-Linguistic Study. In Jeroen van de Weijer (ed.) *Representing phonological detail. Part II: Syllable, Stress, and Sign*. Berlin: Mouton de Gruyter, 171–190.
- Green, Christopher & Stuart Davis (2014). Superadditivity and limitations on syllable complexity in Bambara words. In Ashley Farris-Trimble & Jessica Barlow (eds.) *Perspectives on phonological theory and development, in honor of Daniel A. Dinnsen*. Amsterdam: John Benjamins, 223–247.
- Hayes, Bruce (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.
- Hayes, Bruce, Colin Wilson, & Anne Shisko (2012). Maxent grammars for the metrics of Shakespeare and Milton. *Lg* 88(4), 691–731.
- Hyde, Brett (2007). Non-finality and weight-sensitivity. *Phonology* 24(2), 287–334.
- Hyman, Larry (2006). Word-prosodic typology. *Phonology* 23(2), 225–257.
- Inkelas, Sharon & Draga Zec (1988). Serbo-Croatian pitch accent: the interaction of tone, stress, and intonation. *Lg* 64(2), 227–248.
- Itô, Junko & Armin Mester (1998). *Markedness and word structure: OCP effects in Japanese*. Manuscript, University of California, Santa Cruz.
- Ivić, Pavle (1957). O govoru Galipoljskih Srba. *Srpski dijalektološki zbornik* 12, 1–519.
- Ivić, Pavle (1958). *Die serbokroatischen Dialekte. Ihre Struktur und Entwicklung*. 's-Gravenhage: Mouton.
- Jäger, Gerhard & Anette Rosenbach (2006). The winner takes it all—almost: Cumulativity in grammatical variation. *Linguistics* 44(5), 937–971.
- Jović, Dušan (1968). Trstenički govor. *Srpski dijalektološki zbornik* 17, 1–239.
- Kager, René (1993). The moraic iamb. *CLS* 27, 291–305.
- Kapović, Mate (2015). *Povijest hrvatske akcentuacije: Fonetika*. Zagreb: Matica hrvatska.
- Kawahara, Shigeto (2006). A faithfulness ranking projected from a perceptibility scale: The case of [+ voice] in Japanese. *Lg* 82(3), 536–574.
- Kawahara, Shigeto (2020). A wug-shaped curve in sound symbolism: The case of Japanese Pokémon names. *Phonology* 37(3), 383–418.
- Kawahara, Shigeto & Canaan Breiss (2021). Exploring the nature of cumulativity in sound symbolism: Experimental studies of Pokémonastics with English speakers. *Laboratory Phonology* 12(1).
- Kawahara, Shigeto & Gakuji Kumagai (2021). What voiced obstruents symbolically represent in Japanese: Evidence from the Pokémon universe. *Journal of Japanese Linguistics* 37(1), 3–24.
- Kenstowicz, Michael (2009). Two notes on Kinande vowel harmony. *Language Sciences* 31, 248–270.
- Kim, Seoyoung (2022). A MaxEnt learner for super-additive counting cumulativity. *Glossa: a journal of general linguistics* 7(1).
- Kirchner, Robert (1997). Synchronic chain shifts in Optimality Theory. *LI* 27(2), 341–350.
- Kisseberth, Charles (1970). On the functional unity of phonological rules. *LI* 1(3), 291–306.
- Langston, Keith (1997). Pitch accent in Croatian and Serbian: Towards an autosegmental analysis. *Journal of Slavic linguistics* 5(1), 80–116.
- Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky (1990). Harmonic Grammar—A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, 388–395.
- Lehiste, Ilse & Pavle Ivić (1986). *Word and sentence prosody in Serbocroatian*. Cambridge, MA: MIT Press.
- Ligoris, Orsat (2016). *Serbo-Croatian Accent Retraction: Its Course and Character in the Dialect of Dubrovnik*. PhD dissertation, Leiden University.



- Łubowicz, Anna (2003). *Contrast preservation in phonological mappings*. PhD dissertation, University of Massachusetts Amherst.
- Lyman, Benjamin Smith (1894). The Change from Surd To Sonant in Japanese Compounds. *Oriental Club of Philadelphia*, 1–17.
- Martinović, Martina (2012). The Interaction of Animacy with Phonetic and Phonological Factors in Neoštokavian Pitch Accents. *WCCFL* **29**, 161–168.
- McCarthy, John (2003). Comparative markedness. *Theoretical Linguistics* **29**, 1–51.
- McCarthy, John & Alan Prince (1986). *Prosodic morphology*. Unpublished manuscript, University of Massachusetts Amherst and Brandeis University.
- McCarthy, John & Alan Prince (1994). The emergence of the unmarked: Optimality in prosodic morphology. *NELS* **24**, 333–379.
- McPherson, Laura (2016). Cumulativity and ganging in the tonology of Awa suffixes. *Lg* **92**(1), e38–e66.
- Milenković, Aljoša (2023). Towards a more comprehensive theory of tone-stress interaction. *NELS* **53**(2), 209–218.
- Nishimura, Kohei (2003). *Lyman's Law in loanwords*. Master's thesis, Nagoya University.
- Pater Joe (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**(1), 217–226.
- Pater, Joe (2009b). Weighted constraints in generative linguistics. *Cognitive science* **33**, 999–1035.
- Pater, Joe (2016). Universal grammar with weighted constraints. In John McCarthy & Joe Pater (eds.) *Harmonic Grammar and Harmonic Serialism*. Bristol, CT: Equinox, 1–46.
- Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt & Michael Becker (2010). Harmonic Grammar with linear programming: from linear systems to linguistic typology. *Phonology* **27**(1), 77–117.
- Prince, Alan (1983). Relating to the grid. *LI* **14**(1), 19–100.
- Prince, Alan (1990). Quantitative consequences of rhythmic organization. *CLS* **26**(2), 355–398.
- Prince, Alan & Paul Smolensky (2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Originally circulated as a technical report in 1993. Revised version published by Blackwell, 2004.
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Remetić, Slobodan (1985). Govori centralne Šumadije. *Srpski dijalektološki zbornik* **31**, 1–555.
- Ryan, Kevin (2016). Phonological weight. *Language and Linguistics Compass* **10**, 720–733.
- Ryan, Kevin (2017). Attenuated spreading in Sanskrit retroflex harmony. *LI* **48**(2), 299–340.
- Shih, Stephanie (2017). Constraint conjunction in weighted probabilistic grammar. *Phonology* **34**(2), 243–268.
- Simić, Radoje (1972). Levački govor. *Srpski dijalektološki zbornik* **19**, 1–618.
- Smith, Jennifer (2022). Some Formal Implications of Deletion Saltation. *LI* **53**(4), 852–864.
- Smith, Brian & Joe Pater (2020). French schwa and gradient cumulativity. *Glossa: a journal of general linguistics* **5**(1).
- Smolensky, Paul (1993). *Harmony, markedness, and phonological activity*. Paper presented at Rutgers Optimality Workshop.
- Smolensky, Paul (2006). Optimality in phonology II: harmonic completeness, local constraint conjunction, and feature domain markedness. In Paul Smolensky & Géraldine Legendre (eds.) *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. Vol. 2: Linguistic and philosophical implications. Cambridge, MA: MIT Press, 27–160.
- Staubs, Robert, Michael Becker, Christopher Potts, Patrick Pratt, John McCarthy & Joe Pater (2010). OT-Help 2.0. Software package, University of Massachusetts Amherst. <https://people.umass.edu/othelp/>
- Stevanović, Mihailo (1933). Istočnocrnogorski dijalekat. *Južnoslovenski filolog* **13**, 1–129.
- Tesar, Bruce & Paul Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, MA/London, UK: MIT Press.
- Tomanović, Vaso (1935). Akcenat u govoru Lepetana (Boka Kotorska). *Južnoslovenski filolog* **14**, 59–141.
- Werle, Adam (2009). *Word, phrase, and clitic prosody in Bosnian, Serbian, and Croatian*. PhD dissertation, University of Massachusetts Amherst.
- Yip, Moira (2001). The complex interaction of tone and prominence. *NELS* **31**(2), 531–545.

- Zec, Draga (1999). Footed tones and tonal feet: rhythmic constituency in a pitch-accent language. *Phonology* **16**(2), 225–264.
- Zec, Draga & Elizabeth Zsiga (2010). Interaction of tone and stress in Standard Serbian: Phonological and phonetic evidence. *Formal Approaches to Slavic Linguistics: The Second Cornell Meeting 2009*, 537–555.
- Zec, Draga & Elizabeth Zsiga (2022). Tone and stress as agents of cross-dialectal variation: the case of Serbian. In Haruo Kubozono, Junko Itô & Armin Mester (eds.) *Prosody and Prosodic Interfaces*. Oxford: Oxford University Press, 63–94.
- Zsiga, Elizabeth & Draga Zec (2013). Contextual evidence for the representation of pitch accents in Standard Serbian. *Language and speech* **56**(1), 69–104.
- Zuraw, Kie & Bruce Hayes (2017). Intersecting constraint families: an argument for Harmonic Grammar. *Lg* **93**(3), 497–548.