Towards a more comprehensive theory of tone-stress interaction

Aljoša Milenković

Harvard University

1.   Introduction

In some languages, word-level prominence coexists with tone. In a subset of these languages, prominence and tone crucially interact. While it is widely agreed on that High tone tends towards prosodically strong positions (Goldsmith 1987, Smith 2004, Gordon 2023, among others), the prosodic behavior of High tone in weak positions is less clear. Difficulties in addressing this question arise due to there being divergent approaches to stress. Approaches that analyze stress using metrical constituency (i.e., feet) distinguish between two kinds of weak positions: footed unstressed and unfooted positions, while foot-free approaches make no such distinction. Whether the distribution of High tone in languages that display tone-stress interaction is sensitive to the distinction between footed unstressed and unfooted positions is an outstanding question. Existing Optimality Theoretic (OT; Prince and Smolensky 2004) accounts invariably assume that tone-stress interaction is mediated by metrical structure, thereby predicting such sensitivity. However, these accounts make conflicting predictions. Thus, the central goal of this paper is to systematically compare the predictions of the existing constraint models and test these predictions against some problematic data. In particular, I look into two phenomena that have major implications for the theory of tone-stress interaction: (i) the issue of stress-tone adjacency, whereby word stress and High tone do not coincide but must fall on adjacent metrical units, and (ii) the patterning of contour tones with respect to prominence. I present a revised typology of tone-stress interaction and argue that no existing OT approach successfully generates the full range of cross-linguistic variation. I propose a hybrid model that combines the existing OT models and argue that only this model accommodates all problematic patterns.

2.   Previous accounts: markedness or licensing?

Previous OT accounts of tone-stress interactions are of two kinds. De Lacy (2002) models tone-to-foot mapping via two sets of markedness constraints which penalize Low and Mid tone in foot heads (1a), and High and Mid tone in foot nonheads (1b).

(1)   a.   *HEAD-L/M
            Assess a violation for every instance of Low/Mid tone in the head position of
            a foot.
    b.   *NONHEAD-H/M
            Assess a violation for every instance of High/Mid tone in the nonhead position
            of a foot.

The constraints in 1 form markedness hierarchies such that *HEAD-L universally outranks *HEAD-M and *NONHEAD-H universally outranks *NONHEAD-M.

Per Breteler (2018), tone-stress interaction is regulated by licensing requirements holding between tones and metrical constituents. Licensing requirements hold bidirectionally, given that High tones must be contained within a foot (2a), and feet must must incorporate a High-toned syllable/mora (2b).[1]

(2)   a.   LICENSE(Ft,H), henceforth $\mathscr{L}$(Ft,H)
            Assess a violation for every foot that incorporates no High-toned syllables/moras.
    b.   LICENSE(H,Ft), henceforth $\mathscr{L}$(H,Ft)
            Assess a violation for every High tone associated with no footed syllables/moras.

To date, no attempts have been made to demonstrate the superiority of either of these competing models, although Breteler (2017) provides some cursory remarks. The two approaches crucially differ in terms of headedness. De Lacy (2002)'s markedness-based approach imposes strict headedness requirements on the distribution of tone. *NONHEAD constraints penalize High-toned foot nonheads, but no constraint militates against unfooted Highs. As a corollary, the approach predicts a cross-linguistic dispreference for High-toned nonheads in languages with tone-driven stress: High-toned syllables/moras that are prevented from heading a foot are better off being unfooted than parsed into the weak position of a foot, unless in response to some independent tone-placement constraint (de Lacy 2002: 25–27). In Breteler (2018)'s licensing model, headedness is given no role. $\mathscr{L}$(H,Ft) penalizes unfooted Highs, but no constraint penalizes unstressed footed Highs. Thus, the licensing account treats High-toned foot nonheads as equally unmarked as High-toned foot heads. In what follows, I argue that neither approach's predictions are fully borne out.

3.   Issues

3.1   Stress-tone adjacency: Neoštokavian Serbian

As standardly analyzed (Zec and Zsiga 2010, Zsiga and Zec 2013, Bethin 1998), the standard, Neoštokavian (NS) variety of Serbian has lexical tone which determines the locus of stress. NS allows for at most one singly-linked High tone per word. All non-High-toned syllables are toneless underlyingly and predictably surface with a low pitch.[2] The dialect's

---

[1] See Breteler (2018: 18–19) for the detailed discussion of the LICENSE constraints.

[2] For the present purposes, I disregard the issue of whether the tone-bearing unit in NS is the syllable or the mora (see Zsiga and Zec 2013: 101 for discussion).

stress rule is as follows. Stress regularly falls on the syllable immediately preceding the High-toned syllable (3a). The High-toned syllable receives stress if no syllable precedes (3b). Toneless inputs display initial stress by default and receive a High in the stressed syllable (3c). Thus, NS features both tone-driven stress, given that lexical Highs determine the position of stress (3a–3b), and stress-driven tone, given that the stressed syllable attracts a High in (3c).

(3)   *NS Serbian stress rule*

    a.   Stress the syllable immediately preceding the High-toned syllable.

         ko.ˈšar.káaš   ‘basketball.player.NOM.SG’
         u.sta.ˈno.ví.ti ‘establish.INF’

    b.   Stress the High-toned syllable if no syllable precedes.

         ˈnó.sii.mo   ‘carry.PRS.1PL’   ˈne.nó.sii.mo ‘not.carry.PRS.1PL’
         ˈráa.dii.mo   ‘work.PRS.1PL’   ˈne.ráa.dii.mo ‘not.work.PRS.1PL’

    c.   Stress the initial syllable if no syllable bears a High underlyingly. Insert a High in the stressed syllable.

         ˈpró.pa.de   ‘fail.AOR.2/3SG’   ˈné.pro.pa.de ‘not.fail.AOR.2/3SG’
         ˈpó.ne.se   ‘bring.AOR.2/3SG’   ˈné.po.ne.se ‘not.bring.AOR.2/3SG’

The prosodic behavior of the negation prefix/particle *ne-* in (3b–3c) is indicative of whether a form is toneless or has an initial High underlyingly. In underlyingly toneless forms (3c), High tone alternates between the stem-initial syllable and the negation marker, while forms with a lexical High on the stem-initial syllable (3b) observe no such alternation. Consequently, stress precedes a lexical High in [ˈnenósimo] ‘we do not carry’ (3b) in accordance with (3a). NS stress is quantity-insensitive: generalizations (3a–3c) hold irrespective of the weight of the High-toned syllable and the preceding syllable, and of other syllables in the stress domain.

    The NS data in (3a) illustrate the phenomenon of stress-tone adjacency, whereby the position of stress is determined by lexical tone, but unlike typical cases of tone-driven stress, tone and stress do not coincide. The question arises of what mechanism ensures the observed adjacency effect. On Zec and Zsiga (2010)’s foot-free approach, the adjacency of stress to a lexical High follows from a culminativity constraint which dictates that metrical heads and tonal peaks must be adjacent if unable to coincide. By contrast, Bethin (1998) holds that NS preferentially forms a disyllabic trochee with a High-toned nonhead syllable:

(4)   u.sta.(ˈno.ví.)ti

Restricting stress-tone adjacency to the foot domain, à la Bethin (1998), has two favorable effects. First, the culminativity constraint employed in Zec and Zsiga (2010) to model stress-tone adjacency in NS finds no independent cross-linguistic support. The foot-based approach dispenses with this ad hoc constraint, since the emergence of the ($\sigma\acute{\sigma}$) trochee

as optimal falls out from independently motivated alignment and licensing constraints, as argued below. Second, the foot-based approach is less complex than its foot-free alternative because it generates the NS stress pattern under a single constraint hierarchy, which I demonstrate in Section 4.2.

Assuming that NS exhibits trochaic rhythmic organization, the dialect's stress pattern poses a serious challenge to de Lacy (2002)'s headedness model. Consider the violation profiles of the candidates in tableau (5), which employs the standard *(NON)HEAD constraints, along with ALLFT-L, which is needed to capture default initial prominence in NS.

(5)

	ustanovíti	*HD-L	*NONHD-H	ALLFT-L
a. 💣*	(ˈusta)novíti	*		
b. 💣*	ustano(ˈvíti)			***
c. ☹	usta(ˈnoví)ti	*	*	**

No ranking in (5) favors the intended NS winner over the other two competitors. If ALLFT-L outranks markedness, the winner is the candidate with initial stress, i.e., tone and stress coexist but do not interact. If markedness dominates ALLFT-L, stress invariably falls on the High-toned syllable.

The licensing model readily captures the preference for $(\sigma\acute{\sigma})$ trochees in NS as a joint effect of $\mathscr{L}$(Ft,H) and ALLFT-L:

(6)

	ustanovíti	$\mathscr{L}$(Ft,H)	ALLFT-L
a.	(ˈusta)novíti	*!	
b.	ustano(ˈvíti)		***!
c. ☞	usta(ˈnoví)ti		**

In (6), top-ranking $\mathscr{L}$(Ft,H) ensures that the foot must incorporate a High-toned syllable, while ALLFT-L minimizes the distance between the foot's left edge and the beginning of the stress domain. Of the two candidates that meet $\mathscr{L}$(Ft,H)'s demands (candidates b and c), the trochaic parse with a High-toned syllable in nonhead position (candidate c) fares better on ALLFT-L.

However, the current version of the licensing model is insufficiently restrictive. In underlyingly toneless forms in NS (3c), stress falls on the initial syllable, which receives a High tone. As it stands, the licensing account fails to enforce High tone in the initial, foot-heading syllable, rather than the peninitial syllable, assumed here to constitute the nonhead position of a left-aligned trochee. Both insertion options satisfy $\mathscr{L}$(Ft,H):

(7)

	propade	$\mathscr{L}$(Ft,H)	AllFt-L	Dep-H
a.	('propa)de	*!		
b.	pro('páde)		*!	*
c. ☞	('própa)de			*
d. ☞	('propá)de			*

High tone insertion is driven by $\mathscr{L}$(Ft,H), which requires that the foot be licensed by a High but has no means to determine which of the footed syllables—initial or peninitial—receives a High.

The interim conclusion of this section is that neither existing approach successfully models the complex interaction between lexical tone and prominence in NS Serbian. The theory of tone-stress interaction must be more permissive than the current version of the headedness approach (1) in order to accommodate the preference for ($\sigma\acute{\sigma}$) trochees in NS, and more restrictive than the current version of the licensing approach (2) in order to capture the attraction of Highs to prominent syllables.

3.2 Contour tones and prominence

Previous work on the relationship between tone and prominence has focused predominantly on level tones. The metrical patterning of contour tones has attracted less attention—a notable exception being Qu and Goad (2011). The question of the relationship between contour tones and prominence bears important consequences for the theory of tone-stress interaction, which have yet to be properly assessed.

The prosodic behavior of contour tones lends itself to two possible OT analyses, depending on the underlying assumptions about the representation of contours (see e.g., Yip 2002: 47–52 for discussion). Assuming that contours are separate prosodic units, there are expected to be separate *(NON)HEAD constraints on contour tones: e.g., *HEAD-falling next to the regular *HEAD-L. Analyses along these lines open up a series of questions vis-à-vis ranking: what is the universal markedness scale for contour tones (if any) and how do the *(NON)HEAD constraints on contour tones rank relative to the *(NON)HEAD constraints on level tones?

If contours are represented as sequences of level tones, their behavior must be regulated by the standard *(NON)HEAD constraints of de Lacy (2002). Given that de Lacy (2002)'s *HEAD constraints bar Low and Mid tone from foot-heading syllables, contour tones are treated as marked in foot heads because of the markedness penalty incurred by their Low/Mid components. The presumed markedness of stressed contour tones manifests itself two pathological effects. First, the account predicts languages in which contours repel stress and occur only in non-prominent (particularly unfooted) positions, where they incur no markedness penalties. Second, stressed contours may be subject to elimination, i.e., reduction to a level High.

The markedness of stressed contour tones and resulting repulsion and elimination effects predicted by de Lacy (2002)'s theory are at odds with the typology of contour tone distribution. Notably, the predicted repulsion/avoidance effects are inconsistent with the fact that no language restricts contour tones to unstressed syllables and many languages restrict contour tones to stressed syllables (Zhang 2001).

In conclusion, de Lacy's account predicts unattested patterns, thereby suffering from overgeneration. This section by no means exhausts the problem of the metrical patterning of contour tones. A more detailed exploration of this phenomenon must be left for future work.

4. A midway approach

4.1 Proposal: markedness *and* licensing

In the foregoing, I have pointed out that both approaches discussed herein fail to capture the full range of cross-linguistic variation. De Lacy's markedness-based approach undergenerates, since it strictly prohibits feet with High-toned nonheads. Empirical evidence was brought in from NS Serbian that feet with a High-toned nonhead syllable can be favored over feet headed by a High-toned syllable in languages with tone-driven stress. The markedness-based approach also overgenerates: under certain assumptions, it predicts a tendency to avoid or eliminate contour tones in prosodic heads, contrary to the empirical findings of Zhang (2001). The licensing-based approach fails to capture the attraction of High tone to prominent positions and hence undergenerates.

The only consistent tendency is the unmarked status of High tone in stressed syllables. Even in languages like NS Serbian, where High tone and stress do not regularly coincide, the unmarkedness of stressed Highs emerges in some peripheral interactions in the grammar.[3] The upshot of the present survey is that there is no single universal preference scale for High tone outside stressed positions (contra de Lacy 2002). The present survey identifies two types of interaction:

(8) a. *Type 1*:
         {Stressed Highs, Unfooted Highs} ≫ Footed unstressed Highs
     b. *Type 2*:
         {Stressed Highs, Footed Unstressed Highs} ≫ Unfooted Highs

Type 1 is illustrated by Ayutla Mixtec (Pankratz and Pike 1967, as analyzed in de Lacy 2002): Highs are tolerated in unfooted syllables and barred from foot nonheads. NS Serbian falls into Type 2: Highs in foot nonheads are permitted, whereas unfooted Highs are intolerable. Significantly, the observed variation, both within individual languages and across languages, exceeds the predictions of the markedness-only and licensing-only accounts.

To accommodate the revised typology, I pursue a midway approach which asserts that both headedness and licensing requirements are encoded in the grammar. This hybrid

---

[3]As shown by the data in (3b–3c).

model, sketched in (9), combines Breteler (2018)'s licensing constraints with the *Non-
head markedness constraints of de Lacy (2002).

(9)   a.   *Licensing constraints*:
           $\mathscr{L}$(Ft,H), $\mathscr{L}$(H,Ft)
      b.   *Markedness constraints*:
           *Nonhead-H ≫ *Nonhead-M

In Type 1 languages, *Nonhead dominates licensing, while Type 2 languages observe
the opposite ranking: licensing dominates *Nonhead. Thus far, I have presented *negative*
evidence against both markedness-only and licensing-only models. In what follows, I pro-
vide *positive* evidence for the necessity of both markedness and licensing constraints (in
4.2), and for the elimination of *Head constraints from the system (in 4.3).

4.2   NS stress pattern revisited

The takeaway of Section 3.1 is that NS necessitates both licensing and markedness con-
straints. The licensing account readily explains the regular stress pattern (3a–3b), wherein
stress precedes a lexical High or falls on the High-toned syllable if no syllable precedes.
Markedness is needed to ensure that a High tone is inserted in the stressed rather than the
post-tonic syllable in underlyingly toneless forms (cf. 7).

The constraint model advanced in (9) captures both patterns, as shown in (10). Along
the lines discussed in Section 3.1, $\mathscr{L}$(Ft,H) must outrank AllFt-L. AllFt-L in turn
dominates *Nonhead-H, which explains tone-stress misalignment in (3a).

(10)   a.   *Non-initial lexical High* (3a)

	ustanovíti	$\mathscr{L}$(Ft,H)	AllFt-L	*Nonhd-H
a.	('usta)novíti	*!		
b.	ustano('víti)		***!	
c. ☞	usta('noví)ti		**	*

       b.   *Toneless input* (3c)

	propade	$\mathscr{L}$(Ft,H)	AllFt-L	*Nonhd-H	Dep-H
a.	pro('páde)		*!		*
b.	('propa)de	*!			
c.	('propá)de			*!	*
d. ☞	('própa)de				*

In (10a), the number of AllFt-L violations is minimized at the expense of *Nonhead-
H. In the default pattern (10b), a High tone is inserted in response to top-ranking $\mathscr{L}$(Ft,H),
while the position of an inserted High is determined by lower-ranking *Nonhead-H. This

Aljoša Milenković

constraint interaction instantiates the emergence of the unmarked (McCarthy and Prince 1994). *NONHEAD-H is consistently violated in forms with a noninitial High (10a) and cannot enforce any unfaithful mappings in its favor because it is overridden by faithfulness. However, *NONHEAD-H has the ability to push the inserted High away from the nonhead syllable in forms without a lexical High (10b).

The foot-based analysis outlined in (10) offers a more parsimonious account of the NS facts in (3) than the foot-free analysis of Zec and Zsiga (2010). Zec and Zsiga (2010)'s account, which is couched within Stratal OT (Kiparsky 2000), holds that the insertion of a High tone in underlyingly toneless forms (3c) takes place at the postlexical stratum, primarily because the word-level rankings posited therein cannot accommodate tone-related unfaithful mappings, including the insertion of tone. On the present approach, no appeal is needed to domain stratification, given that the hybrid model captures all relevant generalizations in (3) using a single constraint hierarchy.

4.3 Consequences of the elimination of *HEAD constraints

As argued in Section 3.2, unless the theory admits of a separate markedness scale for contour tones, de Lacy (2002)'s account erroneously restricts contour tones to prosodically weak positions. An explanatory account of the distribution of contour tones must capture their affinity for prominent positions, particularly the fact that contours are restricted to stressed syllables in many languages (Zhang 2001). The hybrid approach in (9) readily solves this problem. The elimination of *HEAD constraints has as a consequence that the Low/Mid portions of contour tones do not incur markedness penalties in prominent positions. Contour tones' affinity for prominent positions follows jointly from the licensing constraints and *NONHEAD-H. $\mathscr{L}$(Ft,H) and $\mathscr{L}$(H,Ft) lure contour tones towards footed positions and away from unfooted positions, respectively. *NONHD-H pushes footed contours with a High-toned component away from the weak position of a foot.

The present account has an added benefit of obviating an appeal to additional markedness constraints introduced in the literature to model certain tone insertion processes. Following Yip (2001), Wuming Zhuang tone sandhi presents a case of stress-driven tone whereby the first of adjacent Low-toned syllables, assumed to head a trochaic foot, receives a High but retains its underlying Low, which results in a falling (HL) contour (underlined in (11) is the presumed foot-heading syllable):

(11)   (L.L) → (HL.L)

This insertion process is problematic for de Lacy-style negative markedness constraints, which penalize stressed Lows but do not explicitly reward Highs in stressed position. Without removing an underlying Low from the stressed syllable, the insertion of a High in foot-heading position is gratuitous, given that the intended winner [HL.L] is equally marked but less faithful than the faithful contender *[L.L]. This is demonstrated in tableau (12).

(12)

σ̀σ̀		MAX-TONE	*HD-L	DEP-TONE
a. 💣*	(σ̱̀σ̀)		*	
b. ☹	(σ̱̂σ̀)		*	*!
c.	(σ̱́σ̀)	*!		

As the driving force behind the problematic tone sandhi pattern in (11), Yip (2001) posits a markedness constraint that penalizes the absence of a High tone in foot heads, dubbed HEAD-H. Alternatively, the insertion process in (11), if in fact motivated by metrical structure, is readily accounted for by the hybrid constraint model this paper argues for. $\mathcal{L}$(Ft,H) drives the insertion of a High in either of the footed syllables, while *NONHEAD-H mandates that a High must not be in the weak syllable of a foot. This is shown in tableau (13).

(13)

σ̀σ̀		$\mathcal{L}$(Ft,H)	*NONHD-H	DEP-H
a.	(σ̱̀σ̀)	*!		
b. ☞	(σ̱̂σ̀)			*
c.	(σ̱̀σ̂)		*!	*

As a consequence of eliminating de Lacy (2002)'s *HEAD constraints from the grammar, Low and Mid tone do not incur a markedness penalty in foot-heading position. This move enables the theory to model some potentially problematic empirical phenomena, including the metrical patterning of contour tones and certain tone insertion processes.

5.   Conclusion

In this paper, I compared the typological predictions of two OT approaches to tone-stress interaction: de Lacy (2002)'s markedness-based account and Breteler (2018)'s licensing-based account. I concluded that neither account generates the full range of cross-linguistic variation on its own. I argued for a constraint model that combines markedness with licensing. This hybrid model was shown to improve the typological coverage of both individual models it combines. Central to the present analysis is the assertion that the markedness of footed unstressed Highs and unfooted Highs is determined on a language-specific basis by the ranking of *NONHD-H relative to the licensing constraints. In languages where markedness outranks licensing (e.g., Ayutla Mixtec), High-toned syllables are either stressed or metrically unparsed but never in the weak position of a foot. In languages where licensing outranks markedness (e.g., NS Serbian), High tone can occur in all footed positions and only in footed positions.

As a follow-up to this preliminary survey, the predictions of the hybrid model proposed herein should be tested against more copious cross-linguistic data.

Aljoša Milenković

References

Bethin, Christina Y. 1998. *Slavic prosody: Language change and phonological theory*. Cambridge University Press.

Breteler, Jeroen. 2017. Deriving bounded tone with layered feet in Harmonic Serialism: The case of Saghala. *Glossa: A journal of general linguistics* 2.

Breteler, Jeroen. 2018. A foot-based typology of tonal reassociation: Perspectives from synchrony and learnability. Doctoral dissertation, University of Amsterdam.

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

Goldsmith, John. 1987. Tone and accent, and getting the two together. In *Annual Meeting of the Berkeley Linguistics Society 13*, 88–104.

Gordon, Matthew. 2023. The phonetic basis for tone-stress interactions: A cross-linguistic study. In *Syllable, stress, and sign*, ed. by Jeroen van de Weijer, 171–190. Walter de Gruyter.

Kiparsky, Paul. 2000. Opacity and cyclicity. *The Linguistic Review* 17:351–367.

McCarthy, John J, and Alan Prince. 1994. The emergence of the unmarked: Optimality in prosodic morphology. In *The Proceedings of NELS 24*, 333–379.

Pankratz, Leo, and Eunice V Pike. 1967. Phonology and morphotonemics of Ayutla Mixtec. *International Journal of American Linguistics* 33:287–299.

Prince, Alan, and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. John Wiley & Sons.

Qu, Chen, and Heather Goad. 2011. On the interaction of tone and stress in contour tone languages: A prosodic account. In *Proceedings of the 2008 Annual Conference of the Canadian Linguistic Association*, 1–39.

Smith, Jennifer L. 2004. *Phonological augmentation in prominent positions*. Routledge.

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

Yip, Moira. 2002. *Tone*. Cambridge University Press.

Zec, Draga, and Elizabeth Zsiga. 2010. Interaction of tone and stress in Standard Serbian: Phonological and phonetic evidence. In *FASL 18: The Cornell Meeting 2008*, ed. by Wayles Browne et al., 536–555. Michigan Slavic Publications.

Zhang, Jie. 2001. The effects of duration and sonority on contour tone distribution: Typological survey and formal analysis. Doctoral dissertation, UCLA.

Zsiga, Elizabeth, and Draga Zec. 2013. Contextual evidence for the representation of pitch accents in Standard Serbian. *Language and speech* 56:69–104.

Aljoša Milenković
aljosamilenkovic@g.harvard.edu