Southern African Linguistics and Applied Language Studies 2014, 32(1): 35–53
Printed in South Africa — All rights reserved

SOUTHERN AFRICAN LINGUISTICS
AND APPLIED LANGUAGE STUDIES
ISSN 1607-3614 EISSN 1727-9461
http://dx.doi.org/10.2989/16073614.2014.925218

Headed tone spans: Binarity and minimal overlap

Micheal Key[1] and Lee Bickmore[2]*
[1]Department of Linguistics, University of Maryland, 1401 Marie Mount Hall, College Park, MD 20742, USA
[2]Department of Anthropology, AS 238, University at Albany, Albany, NY 12222, USA
*Corresponding author, e-mail: L.bickmore@albany.edu

Abstract: We present a theoretical framework for tone that builds on the Headed Spans theory of assimilation. We propose that spans have two important properties, which we term binarity and minimal overlap. Cilungu (Bantu, Zambia) exhibits a process of binary H tone spreading onto the following mora, and then onto the following syllable, unless an Obligatory Contour Principle (OCP) violation would result. We propose two binary constraints on spans to explain this pattern. Further, Cilungu binary spreading obtains regardless of the length of the sequence of H-toned tone-bearing units (TBUs) in the input, which creates arbitrarily long output sequences of H-toned TBUs. We show that the binarity analysis can nonetheless account for this generalisation if spans are permitted to minimally overlap. A welcome consequence of permitting minimal overlap is that the No Crossing Condition can be derived (for tone) from factorial typology: candidates with supra-minimal overlap are harmonically bounded in our theory. Finally, the formal and typological pathologies of non-overlapping alternative analyses are discussed.

Introduction

The study of Bantu tonology has played an important role in the development of phonological theory. In the study of representations, much of the evidence for positing an autonomous tier of features and tones was found in the tonal systems of Bantu languages, and motivated the formulation of autosegmental phonology (Goldsmith 1976a, 1976b). In the development of universal phonological principles, the study of Bantu tonology helped considerably to motivate the Obligatory Contour Principle (OCP) – first as a constraint on underlying representations (Leben 1973), and later as a condition on surface representations, based in no small part on analyses of downstep in Bantu languages (Odden 1986). In the context of Odden's autosegmental framework, he used the observation of adjacent surface Hs in Kishambaa as evidence against a universally inviolable ('strong') OCP. In the development of Optimality Theory (OT) (Prince and Smolensky 1993–2004), the study of Bantu tonology partly inspired a general sub-theory of featural harmony (Cassimjee and Kisseberth 1998) and contributed to the further exploration of OCP effects and the theory of faithfulness (Myers 1997, Bickmore 2000).

Work in OT on featural harmony (Wilson 2003, McCarthy 2004, Smolensky 2006) has shown that several widely adopted OT-based theories of harmony posit a type of 'pro- spreading' markedness constraint that can be satisfied equally well by true harmony as by unattested repairs, such as deletion of blockers to harmony. Theories of harmony with this property include gradient ALIGN (Archangeli and Pulleyblank 1994), local AGREE (Baković 2000), and feature-driven markedness (Beckman 1997, 1998).

In response to these difficulties, McCarthy (2004) proposes a theory of Headed Spans, which makes progress on the 'too many repairs' problem because the pro-harmony constraint it assumes can only be satisfied by minimizing the number of pairs of adjacent feature spans. Given the range of tonal phenomena documented in Bantu languages, it is appropriate to further test the viability of Headed Spans by examining problems in Bantu tonology. The primary goal of this article is to develop Headed Spans into a more general framework in which various cross-linguistic tonal phenomena can be explained. In particular, we motivate two new properties of spans that we argue are crucial to account for tone patterns seen in a variety of Bantu languages: *span binarity* and *minimal overlap*. We propose that span binarity constraints are responsible for a complex pattern of binary H tone spreading in the Zambian Bantu language Cilungu (M.14) (Guthrie 1967–71), as thoroughly described by Bickmore (2007). Binary spreading obtains regardless of the size of the sequence of H-toned TBUs in the input; the generalisation is that a sequence of *n* input H tones maps to an output sequence of *n* + 1 H-toned TBUs; the Cilungu data are presented in sections below.

The proposed span binarity constraints can only account for the generality of Cilungu binary spreading if we further assume that spans may minimally overlap (i.e., coincide on the same TBU). By making overlap a possible representation, we can explain what appears to be a case of '/*n*/ → *n* + 1' H spreading as simple binary /1/ → 2 H spreading because the final two TBUs in a sequence of H-toned TBUs can be parsed into a binary span that overlaps the immediately preceding H span. The proposed constraints demanding binarity and penalizing overlap have an interesting general consequence for the representation of tone: the familiar No Crossing Condition of autosegmental phonology can be derived from constraint interaction rather than from a stipulation (on Gen) in this theory, which we see as an argument in favour of a Headed Spans approach to tone.[1] To demonstrate this, we show that only candidates with minimal overlap or no overlap (of tone spans) are *contenders* (i.e., non-harmonically bounded) in this theory.

Headed Spans

Headed Spans (McCarthy 2004) is a response to certain difficulties encountered by other OT-based theories of harmony, all of which have some sort of 'pro-spreading' constraint (e.g., ALIGN, AGREE). As discussed in Wilson (2003) and McCarthy (2004), the various proposals for the pro-spreading constraint fail for essentially two reasons: (i) the constraint is unable to distinguish between candidates with differing degrees of spreading; (ii) the constraint predicts languages where pseudo-spreading is accomplished by segmental deletion, blocker mutation, selection of shorter allomorphs, affix repositioning, and other unattested repairs.

In response to these unwanted predictions, McCarthy (2004) proposes a theory of *Headed Spans*. A span is a constituent whose peripheral nodes are segments in a contiguous string. There is a span for each distinctive feature or tone, analogous to the tiers of autosegmental phonology. Spans are *headed* by some unique segment belonging to that span, which determines the realisation of all the elements in that span. The idea of headedness harks back to a similar though not identical property of the domains of Optimal Domains Theory (ODT) (Cole and Kisseberth 1995, Cassimjee and Kisseberth 1998).[2] Spans are defined generally by the pair of restrictions on the output of Gen, as in (1).

(1) Restrictions on Gen in Headed Spans
 (a) Segments are exhaustively parsed into spans for each value of each relevant feature or tone.
 (b) Spans have a single head element, which determines the pronunciation of all other elements parsed into the same span.[3]

McCarthy (2004) makes the additional assumption that spans of the same feature value or tone are non-overlapping, although he notes that minimal overlap of the same feature or of distinct tones may prove necessary to analyse contour segments (e.g., affricates) as well as contour tones. We demonstrate below that allowing spans of the same tone to overlap is in fact necessary; we propose that candidates with overlapping spans are necessary to represent the full set of outputs of binary H spreading.

In Headed Spans, spreading is driven by the pressure to minimise the number of adjacent spans of the same feature. This is formalised as a markedness constraint *A-SPAN (F), as defined in (2).

(2) *A-SPAN(F)
 Assign a penalty for each pair of adjacent F spans.

For the tonal phenomena that concern us, we adopt versions of *A-SPAN(F) that apply to spans of H and L in addition to the general version defined in (2). We use T in lieu of F.

(3) (a) *A-SPAN(H)
Assign a penalty for each pair of adjacent H spans.
(b) *A-SPAN(L)
Assign a penalty for each pair of adjacent L spans.
(c) A-SPAN(T)
Assign a penalty for each pair of adjacent tone spans (H or L).

The constraint *A-SPAN(H) (3a) is the Headed Spans equivalent of the constraint OCP(H) in OT analyses of tone that assume autosegmental representations (e.g., Myers 1997, Bickmore 2000). *A-SPAN(H) plays an important role in our analysis due to our assumption that the phonetic contrast between level H tone on consecutive TBUs (HH) and a sequence of a H-toned TBU followed a TBU with a pitch intermediate between H and L (H'H) is expressed in the phonological output, as in (4). We use τ to stand for TBU.

(4) HH *vs*. H'H in Headed Spans
(a) HH (b) H'H
(τ́τ́) (τ́)(τ́)

The assumption about the contrast in realisation between a single H span and adjacent H spans shown in (4) follows the proposal of Bickmore (2000) in that *A-SPAN(H) violations are phonetically interpreted as downstep. The general alternative for the analysis of downstep is to invoke a floating L tone in between the H tones (e.g., Clements and Ford 1979). We don't know of any empirical arguments in favour of this alternative for the analysis of Cilungu. However, Paster and Kim (2011) compare these two approaches and argue that the floating L analysis is required for Tiriki. We concede that it is unclear whether and how output floating elements could be implemented in Headed Spans (McCarthy 2004).

The constraint *A-SPAN(L) compels the minimisation of pairs of adjacent L spans. In most Bantu languages, this constraint seems to be ranked quite low because we nearly always find that H is the active tone. Factorial typology predicts languages in which *A-SPAN(L) is ranked high, and so-called 'tone reversal' Bantu languages show that this pre- diction is borne out.

The *A-SPAN constraints are opposed by the FTHHDSP family of faithfulness con- straints, which replaces earlier IDENT(F) and MAX(F) constraints.[4] The general definition of FTHHDSP is given in (5).

(5) FTHHDSP(T)
If an input TBU $ς_I$ is associated with a tone T and it has an output correspondent $ς_O$, then $ς_O$ is the head of a span of T.

FTHHDSP(T) essentially requires that the output correspondent of the sponsor of a tone head a span of that tone. Looking at things another way, we can say that each feature or tone sponsor is the head of its own span in the input. FTHHDSP(T) simply demands faithfulness to span headedness.

Substituting H and L for T provides us with more specific versions of FFTHHDSP:

(6) (a) FTHHDSP(H)
If an input segment $ς_I$ is associated with an H and it has an output correspondent $ς_O$, then $ς_O$ is the head of an H span.

(b) FTHHDSP(L)
If an input segment $ς_I$ is associated with an L and it has an output correspondent $ς_O$, then $ς_O$ is the head of an L span.

The relative ranking of *A-Span(T) and FthHdSp(L) determines whether a particular language exhibits spreading of H tones or simply maps input L tones faithfully. When *A-Span(T) is ranked above FthHdSp(L), input L tones will be sacrificed under the pressure to minimise adjacent pairs of tone spans, as shown schematically in (7a). When FthHdSp(L) is ranked above *A-Span(T), the pressure to minimise adjacent pairs of tone spans will be suppressed by the need to map input Ls faithfully, as seen in (7b).[5]

(7) (a) *A-Span(T) ≫ FthHdSp(L): Spreading of H tones

/ τ́ τ̀ τ̀ τ̀/	*A-Span(T)	FthHdSp(L)
i. → (τ́ τ́ τ́ τ́)		3
ii. (τ́)(τ̀)(τ̀)(τ̀)	W_3	L

(b) FthHdSp(L) ≫ *A-Span(T): No spreading of H tones

/τ́ τ̀ τ̀ τ̀/	FthHdSp(L)	*A-Span(T)
i. (τ́ τ́ τ́ τ́)	W	L
ii. → (τ́)(τ̀)(τ̀)(τ̀)		3

We assume full specification of tone in the input: each TBU is associated with either H or L. However, many analyses of Bantu tone, both autosegmental as well as OT-based, assume that the input consists of an underlying contrast between H and Ø (Stevick 1969, Hyman and Byarushengo 1984, Myers 1997, Bickmore 2000). In such analyses, L is said to be 'underspecified'. The motivation for this assumption was based on the observation that L tones seem to play no active role in the phonology of most Bantu languages, a fact that can be accounted for by combining underspecification (a H vs. 0/ underlying contrast) with a late default L insertion rule (in rule-based phonology), or an undominated constraint against L tones, *L, in an OT analysis.

There are a number of difficulties with countenancing underspecification in OT. The biggest of these, from the perspective of Headed Spans, concerns the nature of faithfulness. As hinted at from the discussion above, the opposition between *A-Span and FFthHdSp determines whether or not spreading will occur. However, if we assume that inputs do not show a contrast between H and L, the constraint FthHdSp(L) (6b) will be, de facto, inviolable. The consequence of this is that spreading (or strictly speaking, minimisation of adjacent tone spans) will generally be preferred to not spreading, which is clearly not the case in every language. This is not to say spreading would always be *optimal* in an underspecification analysis, because other constraints on spans may still dominate *A-Span(T). The point is rather that spreading must violate a faithfulness constraint that not spreading satisfies. Therefore, we assume that input tone-bearing units (TBUs) contrast between H and L. It should be noted that we are not making a theory-independent argument against underspecification nor are we claiming that all Bantu languages can be successfully analysed by assuming a H vs. L input contrast.[6]

Finally, the constraints governing the position of heads within the span, and thus determining directionality of spreading, are the SpHd variety, shown in (8).

(8) (a) SpHd-{L, R}(H)
The head of a H span is initial/final in that span.
(b) SpHd-{L, R}(L)
The head of a L span is initial/final in that span.

The fact that all H tone spreading in Cilungu is rightward tells us that SpHd-L(H) ≫ SpHd-R(H). There is no evidence to our knowledge support an argument for the relative ranking of SpHd-L(L) and SpHd-R(L).

Span binarity

Bounded vs. unbounded spreading

In Cilungu (Bickmore 2007), there are two distinct types of spreading affecting High (H) tones, both of which are extremely productive.[7] The first is *unbounded spreading*, where an input H is realised not only on its sponsoring TBU, but on an unbounded string (up to the penult of a phrase-final word) of following TBUs as well. The second type of spreading is *bounded spreading*, in which an input H is realised on its sponsoring TBU as well as on a following mora and/or syllable. While the focus of this paper is on the latter process, a brief summary is given here of the environments in which each occurs in Cilungu. In brief, bounded spreading is the 'elsewhere' case in Cilungu: unbounded spreading occurs in the following two specific environments, while all other H tones undergo bounded spreading.[8]

The first case of unbounded spreading concerns the rightmost H tone within the macrostem domain. The *macrostem* in Bantu is the morphological domain which includes the object marker and the following stem. A *stem* includes the root followed by any verbal extension suffixes, followed by the final vowel suffix. This is shown schematically in (9).[9]

(9) Bantu morphological schema
[SM – T/A [$_{macrostem}$ OM [$_{stem}$ Root – Extensions – FV]]]

In Cilungu the OMs, the root, and FV are contrastive for tone. The extensions do not contrast for tone (i.e., they are all L). Thus, the first target of unbounded spreading will be any H on an OM or root as long as it is the rightmost H in the macrostem – that is, as long as the FV is not H-toned. This is illustrated in (10).

(10) Unbounded spreading to the penult of phrase-final word

(a)	/tú-kù-mù-pááràtìk-ìl-à/ 1PL-PRG-3SG-FLATTEN-AP-FV	→	túkúmùpááráátíkílà 'we are flattening for him/her'
(b)	/tú-kù-yá-sùkìlìl-à/ 1PL-PRG-3PL-ACCOMPANY-FV	→	túkúꜜyásúkílílà 'we are accompanying them'
(c)	/tù-ngá-mù-sópòlòl-á/ 1PL-POT-3SG-UNTIE-FV	→	tùngámúꜜsópólòlá 'we can untie him/her'
(d)	/tù-ngá-mù-sùkìlìl-á/ 1PL-POT-3SG-ACCOMPANY-FV	→	tùngámúsùkìlìlá 'we can accompany him/her'

In both (10a) and (10b) the macrostem contains an H (on the root and OM respectively) which is rightmost in its domain and therefore undergoes unbounded spreading. The form in (10c) shows that if the FV is underlyingly H, then a previous H in the macrostem will not undergo unbounded spreading, but rather bounded spreading. The word-initial H on the SM in (10a) and (10b) also undergoes bounded spreading. When that bounded spreading creates a *A-SPAN(H) violation, a phonetic downstep occurs. One final comment about the unbounded spreading in (10a) and (10b) is that, as can be seen, the H does not in fact spread onto the word-final TBU. In Cilungu, unbounded spreading will never penetrate into the final TBU of an intonational phrase. If, however, a word containing a macrostem H is not intonational phrase-final, then the macrostem H will in fact spread to the ultima of the word:

(11) Unbounded spreading of macrostem H to end of non-phrase-final word
/tú-kù-pááràtìk-ìl-à Chòòlà/ → túkúꜜpáápáátíkílá Chòòlà
1PL-PRG-FLATTEN-APP-FV CHOLA 'we are flattening for Chola'

The second instance of unbounded spreading affects a pre-macrostem H which is the rightmost H of the phrase-final word. This is illustrated below.

(12) Unbounded spreading of pre-macrostem H to penult of phrase-final word
/tú-kù-sùkìlìl-à/ → túkúsúkílílà
1PL-PRG-ACCOMPANY-FV
'we are accompanying'

As can be seen, the H on the 1 pl. SM has spread rightward in an unbounded fashion. If, however, the pre-macrostem H is followed by another H in the same word, or is not in the phrase-final word, then no unbounded spreading occurs, as seen in (13).

(13) Unbounded spreading of pre-macrostem H blocked

(a) /tù-ngá-mù-sùkìlìl-á/ → tùngámúsùkìlìlá
1PL-POT-3SG-ACCOMPANY-FV 'we can accompany him/her'

(b) /tú-kù-sùkìlìl-à Chòòlà/ → túkúsùkìlìlà Chòòlà
1PL-PRG-ACCOMPANY-FV CHOLA 'we are accompanying Chola'

Bounded spreading

If an H sponsor is not found in one of the two environments described above, then bounded spreading may occur, depending on whether the following TBU bears a H. If it does, then no spreading of the H sponsored by the first TBU will occur, as there are no available TBUs to spread onto. The result in this case is level H tone on both TBUs, which is usually called *fusion* and is also observed for, e.g., Namwanga (Bickmore 2000). This is shown in (14).

(14) Fusion of consecutive input Hs

(a) /tù-ngá-yá-sópòlòl-á/ → tùngáyásópólòlá 'we can untie them'
1PL-POT-3PL-UNTIE-FV

(b) /tú-á-cí-sópòlòl-á/ → twááćısópólòlá 'we recently untied'
1PL-PST-IMM-UNTIE-FV

The inputs in (14) contain sequences of H-toned sponsors. The outputs with which they are associated contain a sequence of the same number of H-toned TBUs *plus one*, due to the spreading of the sponsored H borne by the rightmost TBU in the sequence. We account for this fact below. For present purposes, the data in (14) serve to demonstrate that not being the rightmost H sponsor in a phrase-final word is a necessary but not sufficient condition for binary spreading; the following syllable must not be associated with a sponsored H.

Bounded spreading as span binarity

If a H sponsor is not the rightmost sponsor in a phrase-final word and the following TBU does not bear a H, then binary spreading is possible, subject to the OCP as discussed below. In the remainder of this section, it will be useful to refer to a *three-syllable window*, which includes the syllable on which some input H is realised and the two syllables immediately following it.

We begin by considering cases in which all syllables in the window are monomoraic. In this case, the sponsoring *mora* heads an H span that is bisyllabic, as shown in (15).

(15) Spreading to following syllable when first two syllables are monomoraic

(a) /sópòlòl-á/ → sópólòlá 'untie!'
UNTIE-FV

(b) /tù-ngá-mù-lém-á/ → tùùngámúꜜlémá 'we can grab him/her'
1PL-POT-3SG-GRAB-FV

In (15b), a bisyllabic H span results which abuts a following H span, which we assume is realised as downstep in Cilungu.[10] That this occurs regardless of the proximity of the following input H shows that the constraint *A-SPAN(H) (the Headed Spans equivalent of the OCP) is violated. We now consider how to account for binary spreading.

Bantu languages have been widely reported to exhibit bounded spreading of H tones. In the vast majority of cases, bounded spreading creates H spans of binary size.[11] How- ever, some apparent cases of binary spreading have been analysed as a phonetic effect of realisation of the *f* 0 peak at the edge of the H-toned syllable (Kim 1999, Myers 1999 for Chichewâ, and Myers 2003 for Kinyarwanda).

Kaplan (2008) has argued that all purported cases of phonological non-iterative tone rules can be analysed as cases of phonetic overshooting of H tone realization. Having considered this possibility for apparent binary spreading in Cilungu, we now present evidence that it is a non-starter. Consider the pair of forms in (16).[12]

(16) Binary spreading is not a late *f*0 peak

(a) /kú-mù-lúàl-ì-à + H/ → kúmúꜜlwáázyá 'and then it healed him/her'
C17-3SG-GE T.WELL-CAUS-FV

(b) /H + kù-mù-lúàl-ì-à + H/ → kúmùlwáázyá 'to heal him/her'
INF-3SG-GET.WELL-CAUS-FV

If binary spreading were reducible to a late *f*0 peak, then we should expect the first two syllables in both forms to be realised with a H tone, resulting in homophony. The floating preprefixal H in (16b) docks on /kù-/ but does not spread onto /mù-/, as it should if late realisation of the *f* 0 peak were the correct analysis. Binary spreading under-applies in (16b), showing that it is sometimes opaque in Cilungu: in common parlance (of rule-based theory), the docking of the preprefixal H *counterfeeds* binary H spreading. Returning to the point, the existence of minimal pairs like the one in (16) clearly show that binary H spreading cannot be a phonetic effect.[13]

The pervasiveness of binary tone spreading in Bantu languages suggests that a constraint must be added in order to distinguish between unbounded and binary spreading. We propose that this is SPBIN(H), as defined in (17).[14]

(17) SPBIN(H)
Assign a violation mark for each H span that does not parse some part (i.e., at least one mora) of exactly two syllables.

SPBIN(H) differs from other binarity constraints proposed in the literature in that it requires *exact* (rather than minimal) binarity. SPBIN(H) is satisfied by candidates with precisely binary spans, whereas minimal binarity constraints like MINIMAL TONE ASSOCIATION (Odden 1998) are satisfied by candidates with H spans that are *at least* binary. SPBIN(H) also differs from tone binarity constraints like BOUND (Myers 1997), which requires that consecutive syllables in a tone span belong to different morphological or prosodic domains. BOUND accounts for the fact that only Hs at the right edge of a domain (e.g., phonological word) undergo binary spreading in Shona. The distribution of binary spreading seen in Cilungu shows that BOUND cannot be a general solution: binary H spreading occurs within domains in this language. In short, SPBIN(H) represents a more general theory of binary spreading because it does not rely on language-particular morphological or prosodic domain restrictions.

The other part of the definition, by which the constraint is satisfied as long as just part of two syllables is in the span, predicts that a H tone will spread onto just the first mora of a following long vowel, rather than onto both morae, which would not satisfy the constraint any better. This is shown below.

By ranking SPBIN(H) above *A-SPAN(T) and FTHHDSP(L), we derive binary spreading:

(18) Bounded spreading as span binarity

/tú-kù-sùkìlìl-à Chòòlà	SPBIN(H)	*A-SPAN(T)	FTHHDSP(L)
(a) → (túkú)(sùkìlìlà Chòòlà)		1	7
(b) (túkú)(sù)(kì)(lì)(là) (Chò)(ò)(là)		W_7	L_1
(c) (túkúsúkílílá Chóólá)	W_1	L	W_8

Tableau (7a) already demonstrated that *A-SPAN(T) ≫ FTHHDSP(L); this ranking favours any degree of spreading over no spreading. Therefore, we can be confident that the unbounded spreading candidate (18c) loses because of the domination of SPBIN(H) over *A-SPAN(T).

Span binarity and *A-SPAN (H)

The fact that binary spreading takes place even when the result is a *A-SPAN(H) violation shows that SPBIN(H) or *A-SPAN(T) ≫ *A-SPAN(H), as in (19).[15]

(19) Spreading at expense of *A-SPAN(H)

/tù-ngá-mù-lém-á/	SPBIN(H)	*A-SPAN(H)	*A-SPAN(T)
(a) → (tù)(ngámú)ꜜ(lém-á)		1	2
(b) (tù)(ngá)(mù)(lém-á)	W_1	L	W_3

When we consider cases in which the first syllable in the window is bimoraic and its first mora is the H sponsor, we find that a trimoraic span is formed, unless the third syllable in the window bears a sponsored H, as in (20).

(20) Bisyllabic spreading blocked by *A-SPAN(H)
(a) /tú-á-mù-lás-á/ → twáámùlásá 'we have hit him/her'
1PL-PST-3SG-HIT-FV
(b) /tú-á-mù-lém-á/ → twáámùlémá 'we have grabbed him/her'
1PL-PST-3SG-GRAB-FV

In (20a-b), the sponsored H on the first syllable spreads onto the second mora of the first syllable, but no further. The reason the H does not spread to the second syllable is because the third syllable bears an underlying H. This would seem to show that *A-SPAN(H) must outrank SPBIN(H) and *A-SPAN(T).

(21) A ranking paradox (cf. (19)): *A-SPAN(H) blocks spreading

/tú-à-mù-lás-á/	*A-SPAN(H)	SPBIN(H)	*A-SPAN(T)
(a) → (twáá)(mù)(lásá)		1	2
(b) (twááamú)ꜜ(lásá)	W_1	L	L_1

At this point, we have paradoxical statements about the ranking of *A-SPAN(H) and SPBIN(H). To resolve this, consider the conditions under which binary spreading either respects or ignores *A-SPAN(H). In (19), when a *A-SPAN(H) violation is not at stake, spreading creates a span that is both bisyllabic and bimoraic. However, in (21) an exclusively bimoraic span is created because spreading further would create an unnecessary *A-SPAN(H) violation. In other words, the demand for bimoraic spreading overrides the preferences of *A-SPAN(H), but *A-SPAN(H) is respected once bimoraic spreading has been achieved. This shows that a single binarity constraint will not suffice. Therefore, we add a second binarity constraint, as in (22).

(22) *MONO-µ (H) (after Cassimjee and Kisseberth 1998, Odden 1998)
Assign a violation mark for every H span that parses exactly one mora.

Taken together, *MONO-µ (H) and SPBIN(H) circumscribe a generalised binary target for H tone spans. Ranked appropriately, *MONO-µ (H) forces minimal spreading (to an adjacent mora), while SPBIN(H) compels spreading to the following syllable. In addition, the definition of SPBIN(H) penalises overspreading (spreading past the second syllable), though it must ultimately be dominated in the overall ranking for Cilungu by the constraint(s) that prefer unbounded spreading in any analysis of the contexts discussed above.

We can account for the fact that bisyllabic spreading fails if the third syllable bears a sponsored H by ranking *A-SPAN(H) above SPBIN(H) and *A-SPAN(T).

(23) *A-Span(H) blocks bisyllabic (supra-bimoraic) spreading

/tú-à-mù-lás-á/	*Mono-μ(H)	*A-Span(H)	SpBin(H)	*A-Span(T)	FthHdSp(H)
(a) →(twáá)(mù)(lásá)			1	2	1
(b) (twááamú)ꜝ(lásá)		W_1	L	L_1	1

Similarly, we can explain why bimoraic spreading obtains regardless of an input H on the third syllable by ranking *Mono-μ (H) above *A-Span(H):

(24) Bimoraic spreading creates *a-span(H) violation

/tú-mù-lém-á/	*Mono-μ (H)	*A-Span(H)	SpBin(H)	*A-Span(T)	FthHdSp(H)
(a)→ (túmú)ꜝ(lémá)		1		1	1
(b) (tú)(mù)(lémá)	W_1	L	W_1	L_2	1
(c) (túmú)ꜝ(lé)ꜝ(má)	W_2	W_2	W_2	L_2	L

Span binarity and mora count

Cilungu does not permit trimoraic syllables. Therefore, considering the first two syllables of the 'window' admits four logical possibilities for moraicity. As shown above, when both syllables are monomoraic, the resulting H span is bisyllabic. We now turn to cases in which the second syllable is bimoraic. In this configuration, the H spreads to either just the first mora of the second syllable or optionally to both morae, unless the second mora of the second syllable bears a H (25b).[16]

(25) Bounded spreading to one or both morae of second syllable

(a) /bélèèng-ìl-á/ → béléèngèlá ~ béléééngèlá 'read for!'
READ-APP-FV

(b) /páàpààtìk-á/ → páápáàtìká ~ páápáátìká 'flatten!'
FLATTEN-FV

(c) /tú-à-zììk-à +H/ → twáázììká 'we buried (far past)'
1PL-PST-BURY-FV

These outcomes are consistent with the rankings argued for in (23) and (24): spreading applies to the next mora, but the presence of an input H-toned mora that is three morae away will (25b) block spreading into the second syllable (*A-Span(H) ≫ SpBin(H)).

In order to account for the optionality in this configuration, SpBin(H) must be such that it is satisfied as long as at least one mora of each syllable is parsed into the H span, else spreading onto both morae of the second syllable, as in (26b), would be incorrectly preferred to spreading to just the first mora. This is shown in the tableau in (26).

(26) Definition of binarity justified

/bélèèng-ìl-á/	*Mono-μ (H)	*A-Span(H)	SpBin(H)	*A-Span(T)	FthHdSp(L)
(a)→ (bélé)(èngè)(lá)	1		1	2	2
(b)→ (béléé)(ngè)(lá)	1		1	2	2

Minimal span overlap

The introduction of the binarity constraints is not alone sufficient to account for all instances of apparent binary H spreading pattern in Cilungu. Assuming that markedness constraints can only evaluate competing output candidates without reference to the input, the addition of binarity constraints alone allows no more than a distinction between spans of binary (bimoraic or bisyllabic) vs. non-binary size.[17] As the data in (27) show, binary spreading appears to create H spans that are not binary. Thus, the labelling of this pattern of spreading as 'binary' is only currently accurate as a rule-like description of the disparity between the number of input Hs and the number of H-toned TBUs in their associated outputs.

(27) Binary spreading creates spans that are more than binary

(a)	/sópòlòl-á/ UNTIE-FV	→	(sópó)(lò)(lá)	'untie!'
(b)	/tù-ngá-sópòlòl-á/ 1PL-POT-UNTIE-FV	→	(tù)(ngásópó)(lò)(lá)	'we can untie'
(c)	/tù-ngá-yá-sópòlòl-á/ 1PL-POT-3PL-UNTIE-FV	→	(tù)(ngáyásópó)(lò)(lá)	'we can untie them'
(d)	/tú-á-ćı-só pòlòl-á/ 1PL-PST-IMM-UNTIE-FV	→	(twááćısópó)(lò)(lá)	'we untied (recent)'
(e)	/tú-á-ćı-yá-sópòlòl-á/ 1PL-PST-IMM-3PL-UNTIE-FV	→	(twááćıyásópó)(lò)(lá)	'we untied them (recent)'

As can be seen from an examination of the forms in (27), each non-phrase-final input sequence of *n* H-toned TBUs maps to an output with an *n* + 1-ary H tone span.

To see that the analysis developed thus far is not able to account for this pattern, consider the violation marks that SPBIN(H) assigns to each of the forms in (27). The output in (27a) receives one violation mark because its final H tone span is unary, while each of the outputs in (27b-e) receive two marks – one for the final H tone span and one for the underlined *n*- ary H tone span (where *n* > 2). Coupled with the fact that *MONO-μ (H) assigns only one mark to each of the candidates in (27), the result is a prediction that binary spreading will not apply to inputs with consecutive Hs. Instead, the current analysis incorrectly predicts that fusion of the adjacent Hs (i.e. no spreading) will be optimal, as illustrated in tableau (28); in fact, candidate (28a) harmonically bounds the desired winner (28c).

(28) Wrong prediction for consecutive input Hs

/tù-ngá-sópòlòl-á/	*MONO-M(H)	*A-SPAN(H)	SPBIN(H)	FTHHDSP(H)	*A-SPAN(T)	FTHHDSP(L)
(a) 💣(tù)(ngásó)(pòlò)(lá)	1		1	1	3	1
(b) (tù)(ngá)ꜜ(sópó)(lò)(lá)	W_2	W_1	W_2	L	W_4	1
(c) →(tù)(ngásópó)(lò)(lá)	1		W_2	1	3	1

If SPBIN(H) and *MONO-μ (H) are to be able explain the binary spreading in (27), we must consider alternative structural descriptions, namely those with overlapping binary spans. If Gen is allowed to produce candidates with overlapping spans, alternative candidates for the surface forms in (27) can compete for optimality. For instance, the final two H-toned TBUs in each underlined span in (27) (i.e., *só pó*) can be parsed into a binary H span which overlaps the H span containing the previous *n* − 1 H-toned TBUs, as illustrated in (29). We henceforth abandon a bracket notation of spans in favour of a tree notation in order to avoid grouping ambiguities.

(29) Binary span overlaps preceding $n - 1$-ary span

(a) /sópòlòl-á/ → sópólòlá ‘untie!’

H_1 L H_2

(b) /tù-ngá-sópòlòl-á/ → tùngásópólòlá ‘we can untie’

L_1 H_1 H_2 L_2 H_3

(c) /tù-ngá-yá-sópòlòl-á/ → tùngáyásópólòlá ‘we can untie them’

L_1 H_1 H_2 L_2 H_3

(d) /tú-á-ćı-sópòlòl-á/ → twááćısópólòlá ‘we untied (recent)’

H_1 H_2 L_1 H_3

(e) /tú-á-ćı-yá-sópòlòl-á/ → twááćıyásópólòlá ‘we untied them (recent)’

H_1 H_2 L H_3

In each output in (29), the output sequence of *n* H-toned TBUs is parsed by a binary span that overlaps a preceding $n - 1$-ary span. In (29a), overlap is vacuous because there are only two H-toned TBUs, but in (29b-e) the binary H span overlaps a preceding $n - 1$-ary H span, rendering the mora of *só* a member of two H spans.

Minimal overlap in Headed Spans reflects proposals for overlapping structure in other domains, including autosegmental structure to represent segmental and tonal contours (Anderson 1976, Goldsmith 1976a, 1976b), ambisyllabicity (Kahn 1976), and foot intersections (as an alternative to unary feet) for parsing the extra syllable in odd-parity forms (Hyde 1999, 2002).

Other languages have binary H spreading that creates output sequences of H-toned TBUs of binary (but not greater) length; examples include Northern Bemba (Sharman and Meeussen 1955) and Ekegusii (Bickmore 1997). Therefore, candidates with overlapping spans must violate some constraint. We propose that this constraint is *OVERLAPSP(H), defined in (30).

(30) *OVERLAPSP(H)

Assign a penalty for each pair of H spans ψ_1, ψ_2 if there is at least one TBU τ such that $\tau \in \psi_1$ and $\tau \in \psi_2$

Notice that the definition of *OVERLAPSP(H) assigns just one mark for each pair of overlapping H spans irrespective of how many TBUs are members of both spans; this is ensured by the language “. . . if there is *at least* one TBU . . . ”. The definition of *OVERLAPSP(H) thus conforms to McCarthy’s (2003) schema for markedness constraints, which requires *categorical* definitions (as in (30)) rather than *gradient* definitions (i.e., ‘Assign one mark [for each pair of overlapping spans [for each TBU that is a member of both spans]]’).[18] We assume that *OVERLAPSP(H) is a member of a family of *OVERLAPSP constraints; for instance, it seems that an analysis of contour tones in this theory would make use of *OVERLAPSP(T), which would penalise pairs of overlapping H and L spans.

The tableau in (31) demonstrates the overlapping analysis of consecutive input Hs under conditions for bounded spreading. Candidate (31b), in which each of the first two input Hs heads its own span, demonstrates that one or more of *MONO-μ (H), *A-SPAN(H), SPBIN(H) AND *A-SPAN(T) must dominate *OVERLAPSP(H). Candidate (31c), which represents fusion but no spreading of the input Hs, shows that FTHHDSP(H) must dominate *OVERLAPSP(H). Finally, while Candidate (31d) is harmonically bounded by (31c), it is shown because it represents the same phonetic form as Candidate (31a) without overlapping structure.

(31) Overlap allows binarity constraints to correctly map consecutive input Hs

/tù-ngá-sópòlòl-á/	*MONO-μ-(H)	*A-SPAN(H)	SPBIN(H)	FTHHDSP(H)	*A-SPAN(T)	FTHHDSP(L)
(a)→ tùngásópólòlá L_1 H_1 H_2 L_2 H_3	1		1	1	3	1
(b) tùngá'sópólòlá L_1 H_1 H_2 L_2 H_3	W_2	W_1	W_2		W_4	L
(c) tùngásópòlòlá L_1 H_1 L_2 H_2	1		1	W_1	3	L
(d) tùngásópólòlá L_1 H_1 L_2 H_2	1		W_2	W_1	3	L

Minimal overlap and the No Crossing Condition

Allowing Gen to produce overlapping span structure requires us to consider a wider range of candidates; the candidates in (29) represent just one possible overlapping analysis – one in which sequences of *n* H-toned TBUs are represented by a binary H span that overlaps a preceding H span of length $n - 1$. An alternative analysis of a sequence of *n* H-toned TBUs consists of $n - 1$ consecutive overlapping binary H spans, as in (32).

(32) $n - 1$ consecutive overlapping binary spans

(a) /sópòlòl-á/ → sópólòlá 'untie!'
H_1 L H_2

(b) /tù-ngá-sópòlòl-á/ → tùngásópólòlá 'we can untie'
L_1 H_1 H_2 L_2 H_3

(c) /tù-ngá-yá-sópòlòl-á/ → tùngáyásópólòlá 'we can untie them'
L_1 H_1 H_2 H_3 L_2 H_4

(d) /tú-á-ćí-sópòlòl-á/ → twááćísópólòlá 'we untied (recent)'
H_1 H_2 H_3 H_4 L H_4

(e) /tú-á-ćí-yá-sópòlòl-á/ → twááćíyásópólòlá 'we untied (recent)'
H_1 H_2 H_3 H_4 H_4 L H_6

Although the candidates in (32) receive the same phonetic interpretation as those in (29), they differ in their constraint violation profiles. We now compare the two sets of candidates, as well as two other logically possible types of overlap in order to answer the question of whether any classes of overlapping structures are harmonically bounded.

Since harmonic bounding entails impossibility under a particular hypothesis about CON, we seek an explanation for any substantive universals regarding the manner in which tone spans may be assigned to TBUs. The other way of expressing universals in OT is to impose a stipulation on Gen: if a particular kind of structure is universally unattested, it can be ruled out by fiat in the statement of Gen. However, as has been argued for since Prince and Smolensky (1993/2004), a responsible definition of Gen rules out only those structures which have no meaningful linguistic interpretation – e.g., syllables with feet as dependents. Since CON contains all the substantive statements about structures, any universals that can be explained by virtue of CON alone add to the explanatory capacity of OT. Thus, we seek to investigate the question of the typology of tone span-to-TBU relations in terms of Con, rather than by making a stipulation on Gen. This is carried out in the violation (non- ranking) tableau in (33) by considering an input with a sequence of an arbitrary number of H-toned TBUs (which are in the appropriate context for binary spreading, denoted as τ̀. . . τ́).

(33) Supraminimal overlap harmonically bounded

/ τ́ τ́ τ́ τ̀ …τ́/	*OverlapSp(H)	FthHdSp(H)	SpBin(H)
(a) τ́ τ́ τ́ τ́ H_1 H_2 H_3	2		
(b) τ́ τ́ τ́ τ́ H_1 H_2	1	1	1
(c) τ́ τ́ τ́ τ́ H_1 H_2	1	1	2
(d) τ́ τ́ τ́ τ́ H_1 H_2 H_3	3		1

Candidates (33c, d) are both harmonically bounded. Candidate (33c), which contains a pair of overlapping ternary spans, is harmonically bounded by candidate (33b) because the former incurs an additional violation of SpBin(H), while the other constraints do not distinguish the two.[19] Candidate (33d), which contains three consecutive overlapping spans (one ternary, two binary), is harmonically bounded by (33a), which performs better on *OverlapSp(H) and SpBin(H).

These two observations point to an important conclusion about the set of *overlapping* structures emitted by Gen for an input containing *n* consecutive Hs: only candidates with $n - 1$ consecutive overlapping binary spans or candidates with an *n*-ary span that overlaps a binary span are optimal under some ranking of the constraints. This result is a consequence of constraints demanding that H spans be of binary size (*Mono-μ(H) and SpBin(H)) and a constraint penalising each pair of overlapping spans. Thus, at least in the case of tone, we derive the equivalent of the No Crossing Condition of autosegmental phonology (Clements and Ford 1979, Goldsmith 1976a, 1976b, Hammond 1988, Sagey 1988) from an independent hypothesis about Con; no restriction on the output of Gen is required.

One further question regarding overlapping spans is how precedence relations are computed. We have shown above that only candidates with *minimal* overlap (i.e., only a single TBU is a member of two spans) are non-harmonically bounded. Due to this fact, we argue that temporal relations can be straightforwardly derived from the ordering of spans. If spans are interpreted as sets of articulatory instructions, as was proposed for autosegmental representations (Hammond 1988), then each tone issues an instruction to one or more TBUs (but not vice-versa). On this interpretation, the issue of precedence relations is trivial: we can simply assume that the set of instructions issued by the first span strictly precedes the set of instructions issued by the second span.

Note that overlapping H spans do not violate *A-Span(H) because the span edges are not actually adjacent, but rather coincide on the same TBU. We choose to make explicit the notion of adjacency by adding a second clause to the definition of *A-Span(H), as in (34), to ensure that it will only be violated at loci of true adjacency (realised as downstep), but not by a pair of overlapping H spans (realised as level H tone).

(34) *A-Span(H) (revised)
Assign a penalty for each pair of adjacent H spans ψ_1, ψ_2 if there is no element ε such that $\varepsilon \in \psi_1$ and $\varepsilon \in \psi_2$.

Note that the addition of the clause pertaining to overlap simply details what is not included in the notion of adjacency that the *A-Span family makes use of.

Now we restrict our attention just to candidates (33a-b). Candidate (33a), which contains three binary spans, betters candidate (33b) on both FthHdSp(H) and SpBin(H), but is worse on *OverlapSp(H). This means candidate (33b) can only be optimal in a constraint hierarchy in which *OverlapSp(H) dominates both FthHdSp(H) and SpBin(H), since these are the only other constraints that distinguish the two candidates. While such a ranking favours (33b) over all more gratuitously *overlapping* competitors, the optimal candidate under the same ranking would be one with no overlap at all. Therefore, candidate (33b) is collectively harmonically-bounded (Samek-Lodovici, Vieri and Prince 2002) by overlapping candidates with consecutive binary spans, and by candidates with no overlap.

In the theory presented here, the choice of analysis of this pattern comes down to $n - 1$ consecutive overlapping binary H spans versus a single H span. However, since both types of candidates always have the same phonetic realisation, an independent argument in favour of the overlap analysis is required. We provide this in the following subsection.

An alternative: Non-overlapping analyses

In this subsection, we consider an alternative analysis of Cilungu that does not rely on *Mono-μ(H), SpBin(H) and *OverlapSp(H). On this view, we assume that Gen does not produce candidates with overlapping span structure, and thus some other constraint(s) must prefer the mapping from n input Hs to a single H span with $n + 1$ H-toned TBUs over mapping to candidates with spans of $n + 2$ (or more) H-toned TBUs. We show below that two versions of this alternative non-overlapping analysis require one or both of the following assumptions: (i) the number of constraints in Con is unboundedly large; (ii) Con contains anti-faithfulness constraints (Alderete 2001). These assumptions are either psychologically (i) or theoretically (in OT) (ii) untenable.

The theory of tone spans we have proposed above permits two fundamental types of candidate representations for the outputs of the Cilungu /n(H)/ → n + 1(τ́) mappings: candidates with overlapping spans and candidates with n + 1-ary spans. This is a consequence of the proposed contrast in phonetic interpretation between a single span or overlapping spans, which are both realised with level tone, and strictly adjacent spans of the same tone value, which are realised as downstep.

Candidates with overlapping spans and candidates without overlapping spans differ in the type of analysis given to the /n(H)/ → n + 1(τ́) mappings. As demonstrated above, candidates with overlapping spans give an analysis of this pattern as 'binary spreading', due to the activity of SpBin(H) in selecting all candidates with $n + 1$ H-toned TBUs over candidates with $n + 2$ or greater. In contrast, candidates without overlapping spans represent the /n(H)/ → n + 1(τ́) mappings with a n + 1-ary span. This alternative analysis of these mappings can be thought of as an *endless chain shift*, as it is dubbed by Prince (2007), also known as *unconditional augmentation* (McCarthy 2002, Moreton 2004). An endless chain shift is defined in (35).

(35) Definition of an endless chain shift
Let A, B, C . . . be a subset *in* ⊆ In of the set of inputs to the grammar. A function F defined on *in*: </A/,[B]>,</B/,[C]>,</C/,[D]>. . . } is an endless chain shift (in a particular language) if there is no ordered pair hx, yi∈ F such that x=y.

The definition in (35) describes a chain shift in which /A/ → [B], /B/ → [C], /C/ → [D] . . . , with no terminal idempotent mapping /X/ → [X]. This function is unlike familiar types, such as a vowel chain shift in which input low vowels raise to mid, input mid vowels raise to high, and input high vowels are mapped faithfully. Under the non-overlapping analysis, we know that Cilungu maps n input Hs to a single span of $n + 1$ H-toned TBUs for arbitrarily large values of n (constrained only by the morphotactics), so we cannot discern any terminal idempotent mapping of n input Hs.

In order to analyse Cilungu as an endless chain shift, some constraint(s) other than SpBin(H) must invariably prefer candidates with a H span of size $n + 1$ to those with a H span of size n. We consider two versions of this analysis.

One version proposes a family of arbitrary span size constraints preferring spans of each size to those of other sizes (i.e., SpTern(H), SpQuat(H), SpQuin(H) etc.). This would require not only that each of these constraints be ranked above *A-Span(T), but for all constraints in the putative SP-*n*(H) family, SP-*n*(H) $\gg$ SP-*n* − 1(H). This is only possible if we also introduce 'retentive' faithfulness constraints, such as locally-conjoined faithfulness constraints (Kirchner 1996), which *do not* penalise a mapping from an input with a sequence of *n* Hs onto a candidate with a span of *n* + 1 Hs, but assign a penalty to candidates with H spans of size *n* + 2 or greater. Although the series of SP-*n*(H) constraints form a harmonic scale (Prince and Smolensky 1993–2004), the scale formed in this case is quite odd because there is no least marked state. Since *n* Hs map onto H spans of size *n* + 1 for arbitrarily large values of *n*, this harmonic scale must be unbounded in size. This implies that Con itself is unbounded in size, if individual constraints are counted, which is not psychologically plausible.

A second version of this analysis proposes a single markedness constraint like Extend-R(H) (Cassimjee 1995, Bickmore 2000), defined in (36).

(36) Extend-R(H)
Let τ_I be the rightmost H-toned TBU in the input and let τ_O be its output correspondent. Assign a penalty if τ_O is the final TBU in the H span.

The effect of Extend-R(H) is that is always favors rightward spreading. In order for this approach to work, we again require a retentive faithfulness constraint that doesn't penalise a mapping from an input with a sequence of *n* Hs onto a candidate with a span of *n* + 1 Hs, but assigns a penalty to candidates with H spans of size *n* + 2 or greater. Extend-R(H) is neither a markedness constraint nor a faithfulness constraint, but rather a third type known as *anti-faithfulness* (Alderete 2001), which *requires* an unfaithful mapping in order to be satisfied. Anti-faithfulness constraints have only been proposed to exist in the output-output correspondence dimension, and there are several reasons for excluding them from the input-output dimension; among them is the dubious prediction (under factorial typology) of languages with purely phonological exchange processes (Moreton 2004).

We have demonstrated the failure of two non-overlapping analyses of the /*n*(H)/ → *n* + 1($\acute{\tau}$) mappings. While it is possible to imagine other versions of this type of analysis, the condition that a single H span of *n* + 1 TBUs must always be preferred to a span of *n* TBUs means that other possible analyses of this kind must still countenance an endless chain shift in some form. Assuming that there are no true cases of endless chain shifts in natural languages (McCarthy 2002, Moreton 2004), we take the difficulties facing any such analysis of the Cilungu data as independent motivation for our analysis.

Conclusion

In this article, we have argued for an extension of the theory of Headed Spans (McCarthy 2004) to account for a complicated pattern of bounded H tone spreading facts in Cilungu. We have shown that bounded spreading is phonological rather than phonetic in this language. Assuming an underlying contrast between H and L, we have argued for an analysis in which bounded spreading is driven by a constraint that demands than H spans parse more than one mora, and by a constraint demanding H spans that parse morae belonging to two syllables. The evidence that these are separate constraints is the activity of the OCP in blocking only supra-bimoraic spreading.

Based on our assumption that a H$^{!}$H sequence is parsed as a pair of adjacent H spans we argued that the same bounded spreading process that creates binary H-toned TBU sequences also creates sequences of unbounded length. In order to account for this fact using just the proposed binarity constraints, we proposed that H spans must be permitted to overlap. We showed that a constraint against overlapping spans restricts the class of overlapping structures to those in which the degree of overlap is minimal; candidates with supraminimal overlap were shown to be harmonically bounded and therefore the No Crossing Condition of autosegmental phonology was derived from our theory of CON, rather than by the traditional (often implicit) restriction on Gen. Finally, we showed the superiority of the overlapping analysis of Cilungu by showing that two versions of a non-overlapping (endless chain shift) analysis require an implausible psychological reality (an unbounded number of constraints) or predict phonological patterns that have never been observed.

The principal avenue for future research in this program concerns the predicted typology of the binarity and overlap constraints. For instance, factorial typology predicts a language in which the binarity constraints dominate the constraints on span directionality. In such a language, spreading could reverse direction when faced by word boundaries or neighbouring tones (depending on the disposition of other constraints). Marlo (2007: 170–172) describes a particular tone pattern of Lumarachi (J.30) in which a melodic H docks onto first mora of the second syllable of a trisyllabic (or longer) stem and then spreads one mora to the right. However, in CVCV stems, the melodic H docks on to the final vowel and spreads one mora to the left.

The other avenue for future work lies in distinguishing the predictions of Headed Spans as a function of the relationship between the Gen and Eval components of OT. Our analysis assumes the classic parallel evaluation originally argued for in Prince and Smolensky (1993–2004). In parallel OT, Gen is unrestricted in its ability to construct candidates from a typically tacit universal phonological alphabet, which produces a denumerably infinite set of candidates. Eval then assigns violation marks to the entire set of candidates in one pass. Alternatively, the Gen-Eval relationship could be serial, according to which Gen would be restricted to some set of candidate-building operations (e.g., one instance of epenthesis). The result of this is a finite and often small candidate set. Since the derivation of input forms sometimes requires more than one operation to apply before the output form is reached, serial OT must assume that there is a feedback loop from Eval to Gen such that the evaluation of the candidate set on pass *n* serves as the input to Gen on pass *n* + 1.

There is a growing literature that directly examines the predicted typologies of parallel vs. serial OT with respect to various phenomena, including cluster simplification (McCarthy 2008), opacity (McCarthy 2007), metrical stress assignment (Pruitt 2010), optionality (Kimper 2011), and most relevant to present concerns, harmony (McCarthy 2010).[20] It would thus be interesting to compare the predictions of the present approach, based in parallel OT, with a serial version of Headed Spans or the serial theory of harmony presented in McCarthy (2010).

Notes

1 For analogous attempts to derive the No Crossing Condition in rule-based autosegmental phonology, see Hammond (1988) and Sagey (1988).

2 One difference between Headed Spans and ODT is whether the head of a span must actually be a member of the span: in Headed Spans this is obligatory (stipulated in the definition of Gen), while it is expressed as a violable constraint in ODT. Another difference is expressed in (1b). It should be noted that we make no attempt in this paper to argue for Headed Spans over ODT.

3 Key (2007) and O'Keefe (2007) propose that (1b) is too strong. Key (2007) analyses tone displacement as spreading plus non-realisation of the tone by the span head, while O'Keefe's evidence comes from cases of transparency in vowel harmony systems.

4 See McCarthy (2004) for arguments in favour of FthHdSp over traditional faithfulness constraints.

5 All ranking tableaux in this paper are presented in the comparative format (Prince 2002) rather than the classic violation style.

6 See Prince (2007) for an analogous discussion of the history and confusion over claims about constraint violability.

7 The Cilungu data in this paper were taken from Bickmore (2007).

8 The discussion here is confined to H tones which are underlyingly linked to some root or affix. Cilungu, like a number of other Bantu languages, also exhibits a grammatical or suffixal H tone which is present in certain tenses/aspects – its highly morphologically-conditioned patterns of docking and spreading are detailed in Bickmore (2007) but will not concern us here.

9 SM = subject marker, T/A = tense/aspect marker, OM = object marker, FV = final vowel.

10 Headed Spans dispenses with output floating elements, and thus theories of downstep that rely on floating L tones in between Hs cannot be maintained in Headed Spans.

[11] A notable exception is a case of ternary H spreading in the Zezuru dialect of Shona reported in Myers (1987).

[12] The '+H' in the input in (16) signals the presence of a melodic H, which appears in certain tenses/ aspects/moods and is ultimately realised on the second and subsequent stem morae (and fuses with a stem-initial H if one is present). The 'H+' is a floating tone prefix, which is an allomorph of the preprefix /ú -/.

[13] Of course, late f0 realisation may still serve as a plausible diachronic mechanism for the existence of phonological binary H spreading in Cilungu.

[14] There are alternative solutions involving faithfulness or anti-faithfulness. An example of the latter is briefly discussed in the article. The former would analyse binary spreading with the addition of a faithfulness constraint that penalises spreading only if it is greater than one syllable or mora. An example of this type of analysis is Kirchner's (1996) use of local conjunction of faithfulness constraints to explain chain shifts. However, see, e.g., McCarthy (2007) for a discussion of local conjunction and its typological difficulties. Since the data presented here do not decide between the competing analyses, we will not discuss this issue further.

[15] Despite their definitions, *A-SPAN(H) and *A-SPAN(T) do not stand in a *stringency relation* because they disagree in their preferences for actual candidates; Prince (1997) called this situation 'phony stringency'. Therefore, they can be ranked by a direct argument.

[16] As explained in connection with (16) above, '+H' represents a melodic H tone which marks this tense/aspect, and which docks onto the second and subsequent morae of the *stem*. As in the example in (16), the tone pat- tern seen in (25b) shows that docking of the melodic H applies before the application of binary H spreading, which spreads the input H of /tú -/ to the following mora, but not to the following syllable, due to the blocking effect of *A-SPAN(H) (i.e., because the melodic H has docked on the second mora of the stem), as discussed.

[17] This fact is a consequence of the parallel architecture of OT, and of the fact that classic markedness constraints do not refer to input representations. If Gen and Eval were looped and Gen were restrained to emitting candidates that are minimally unfaithful to the input, as in persistent OT (McCarthy 2006) and OT-CC (McCarthy 2007), then the binarity constraints could be sufficient.

[18] Categorical constraints contain no more than one universal quantifier in their definition.

[19] FTHHDSP(L) has been omitted from the tableau because it does not distinguish overlapping structures.

[20] A very useful classified bibliography of work to date in serial OT has been assembled by John McCarthy. It can be accessed at http://works.bepress.com/john j mccarthy/102/

References

Alderete J. 2001. Dominance effects as transderivational anti-faithfulness. *Phonology* 18: 201–253.

Anderson SR. 1976. Nasal consonants and the internal structure of segments. *Language* 52: 326–344.

Archangeli D, Pulleyblank D. 1994. *Grounded phonology*. Cambridge, MA: MIT Press.

Baković E. 2000. Harmony, dominance, and control. PhD thesis, Rutgers University, New Brunswick, NJ.

Beckman J. 1997. Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* 14: 1–46.

Beckman J. 1998. Positional Faithfulness. PhD thesis, University of Massachusetts, Amherst, MA.

Bickmore L. 1997. Problems in constraining high tone spread in Ekegusii. *Lingua* 102: 265–290.

Bickmore L. 2000. Downstep and fusion in Namwanga. *Phonology* 17: 297–331.

Bickmore L. 2007. *Cilungu phonology*. Stanford, CA: CSLI.

Cassimjee F. 1995. Isixhosa tonology: an optimal domains theory analysis. Ms., University of Illinois, Urbana-Champaign, IL.

Cassimjee F, Kisseberth C. 1998. Optimal Domains Theory and Bantu tonology: a case study from Isixhosa and Shingzidja. In: Hyman L, Kisseberth C (eds), *Theoretical aspects of Bantu tone*. Stanford, CA: CSLI. pp 33–132.

Clements, GN, Ford KC. 1979. Kikuyu tone shift and its synchronic consequences. *Linguistic Inquiry* 10: 95–108.

Cole J, Kisseberth C. 1995. An Optimal Domains Theory of harmony. Ms, University of Illinois. Urbana-Champaign, IL.

Goldsmith J. 1976a. *Autosegmental phonology*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.

Goldsmith J. 1976b. An overview of autosegmental phonology. *Linguistic Analysis* 2: 23–68.

Guthrie M. 1967–71. *Comparative Bantu*. Farnborough: Gregg. 4 vols.

Hammond M. 1988. On deriving the Well-Formedness Condition. *Linguistic Inquiry* 19: 319–325.

Hyde BS. 1999. Overlapping feet in Polish. In: Dziwirek K, Coats H, Vakareliyska C (eds), *Formal approaches to Slavic linguistics VII*: 185–203.

Hyde BS. 2002. A restrictive theory of metrical stress. *Phonology* 19: 313–359.

Hyman LM, Byarushengo ER. 1984. A model of Haya tonology. In: Clements GN, Goldsmith J (eds), *Autosegmental studies in Bantu tone*. Dordrecht: Foris. pp 53–103.

Kahn D. 1980. *Syllable-based generalizations in English phonology*. New York, Garland.

Kaplan AF. 2008. Peak delay and tonal noniterativity. Available at http://roa.rutgers.edu/files/972-0508/972-KAPLAN-0-0.PDF [accessed 5 May 2014].

Key M. 2007. Headed Spans and Bantu tonology. In: Bateman L. (ed.), *University of Massachusetts occasional papers in linguistics: papers in Optimality Theory III, volume 32*. Amherst, MA: GLSA. pp 187–207.

Kim SA. 1999. Positional effect on tonal alternation in Chichewˆ a: phonological rule vs. phonetic timing. *Proceedings of the Chicago Linguistic Society* 34: 245–257.

Kimper W. 2011. Locality and globality in phonological variation. *Natural Language and Linguistic Theory* 29: 423–465.

Kirchner R. 1996. Synchronic chain shifts as local conjunction. *Linguistic Inquiry* 27: 341–350.

Leben W. 1973. Suprasegmental phonology. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.

Marlo M. 2007. The verbal tonology of Lunyala and Lumarachi: two dialects of Luluyia (Bantu, J.30, Kenya). PhD thesis, University of Michigan, Ann Arbor, MI.

McCarthy JJ. 2002. *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.

McCarthy JJ. 2003. OT constraints are categorical. *Phonology* 20: 75–138.

McCarthy JJ. 2004. Headed Spans and autosegmental spreading. Available at http://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1041&context=linguist_faculty_pubs. [accessed 5 May 2014].

McCarthy JJ. 2006. Restraint of analysis. In: Baković E, Ito J, McCarty JJ (eds), *Wondering at the natural fecundity of things: essays in honor of Alan Prince*. University of California, Santa Cruz, CA: Linguistic Research Center.

McCarthy JJ. 2007. *Hidden generalizations: phonological opacity in Optimality Theory*. London: Equinox.

McCarthy JJ. 2008. The gradual path to cluster simplification. *Phonology* 25: 271–319.

McCarthy JJ. 2010. Autosegmental spreading in Optimality Theory. In: Hume E, Goldsmith J, Wetzels L (eds), *Tones and features*. Berlin: Mouton de Gruyter. pp 195–222.

Moreton E. 2004. Non-computable functions in Optimality Theory. In: McCarthy JJ (ed.), *Optimality Theory in phonology*. Oxford: Blackwell. pp 141–163.

Myers S. 1987. Tone and the structure of words in Shona. PhD thesis, University of Massachusetts, Amherst, MA.

Myers S. 1997. OCP effects in Optimality Theory. *NLLT* 15: 847–892.

Myers S. 1999. Tone association and f0 timing in Chichewˆ a. *Studies in African Linguistics* 28: 215–239.

Myers S. 2003. F0 timing in Kinyarwanda. *Phonetica* 60: 71–97.

Odden D. 1986. On the role of the Obligatory Contour Principle in phonological theory. *Language* 62: 353–383.

Odden D. 1998. Principles of tone assignment in Tanzanian Yao. In: Hyman L, Kisseberth C (eds), *Theoretical aspects of Bantu tone*. Stanford, CA: CSLI. pp 265–314.
O'Keefe M. 2007. Transparency in Span Theory. In: Bateman L, O'Keefe M, Reilly E, Werle A (eds), *University of Massachusetts occasional papers in linguistics 32: papers in Optimality Theory III*. Amherst, MA: GLSA. pp 239–258.
Paster M, Kim Y. 2011. Downstep in Tiriki. *Linguistic Discovery* 9: 71–104.
Prince A. 1997. Stringency and anti-Paninian rankings. *Handout from the LSA Institute*, Cornell University, Ithaca, NY.
Prince A. 2002. Arguing optimality. In: Carpenter A, Coetzee A, de Lacy P (eds), *University of Massachusetts occasional papers in linguistics 26: papers in Optimality Theory II*. Amherst, MA: GLSA. pp 269–304.
Prince A. 2007. The pursuit of theory. In de Lacy P (ed.), *The Cambridge handbook of phonology*. Cambridge: Cambridge University Press. pp 22–46.
Prince A, Smolensky P. 1993–2004. *Optimality Theory: constraint interaction in generative grammar*. Oxford: Blackwell.
Pruitt K. 2010. Serialism and locality in constraint-based metrical parsing. *Phonology* 27: 481–526.
Sagey E. 1988. On the ill-formedness of crossing association lines. *Linguistic Inquiry* 19: 109–118.
Samek-Lodovici V, Prince A. 2002. Fundamental properties of harmonic bounding. Report No. RuCCS-TR-71. New Brunswick, NJ: Rutgers Center for Cognitive Science, Rutgers University.
Sharman JC, Meeussen AE. 1955. The representation of structural tones, with special reference to the tonal behaviour of the verb, in Bemba, Northern Rhodesia. *Africa* 25: 393–404.
Smolensky P. 2006. Optimality in phonology II: harmonic completeness, local constraint conjunction, and feature-domain markedness. In: Smolensky P, Legendre G (eds), *The harmonic mind, volume 2*. Cambridge, MA: MIT Press. pp 27–160.
Stevick E. 1969. *Tone in Bantu*. *IJAL* 35: 330–341.
Wilson C. 2003. Unbounded spreading in OT (or, Unbounded spreading is local spreading iterated unboundedly). Handout of paper presented at Southwest Optimality Theory Workshop, University of Arizona, Tucson.