Sequences of high tones across word boundaries in Tswana

Sabine Zerbian

University of Stuttgart

sabine.zerbian@ifla.uni-stuttgart.de

Frank Kügler ⓘD

Goethe-Universität Frankfurt

kuegler@em.uni-frankfurt.de

The article analyses violations of the Obligatory Contour Principle (OCP) above the word level in Tswana, a Southern Bantu language, by investigating the realization of adjacent lexical high tones across word boundaries. The results show that across word boundaries downstep (i.e. a lowering of the second in a series of adjacent high tones) only takes place within a phonological phrase. A phonological phrase break blocks downstep, even when the necessary tonal configuration is met. A phrase-based account is adopted in order to account for the occurrence of downstep. Our study confirms a pattern previously reported for the closely related language Southern Sotho and provides controlled, empirical data from Tswana, based on read speech of twelve speakers which has been analysed auditorily by two annotators as well as acoustically.

1 Introduction

Although some Bantu languages have lost tone (most notably Swahili), the vast majority of Bantu languages are two-tone languages with a surface high (H) and low (L) level tone. Tone in Bantu languages has a lexical and a grammatical function (see Kisseberth & Odden 2003, Downing 2011, Marlo & Odden 2019 for overviews). Phonologically, in many Bantu languages only lexical high tones are assumed to be underlyingly represented, and low tones are inserted late in the derivation as default tones (Hyman 2001; Yip 2002; Marlo & Odden 2019: 151–153). High tones can be observed to be the active tones, taking part in tone shift, spread, deletion, and/or fusion. Some examples will be provided below. Bantu languages are particularly known for their tonal mobility (see Yip 2002: 66) and tone sandhi. Tone sandhi, i.e. tonal changes due to neighbouring tones, can be observed at the word and phrase level. The focus of the article is on the phonetic lowering of a lexical high tone due to the presence of a preceding high tone. This process is commonly referred to as downstep. Two adjacent high tones constitute a typical environment which violates the Obligatory Contour Principle (OCP; Leben 1973) prohibiting identical adjacent elements. Although they violate the OCP they constitute the context for downstep. This study addresses the conditions for downstep in Tswana analysing it as a phrasal phenomenon.

   Our study has the aim of presenting an empirical study for establishing the structural conditions for downstep in (a variety of) Tswana, taking both number of adjacent high tones and

First published online 6 October 2021

syntactic structure into consideration as suggested in previous literature. Next to a qualitative analysis based on auditory impression, the current study also provides a quantitative acoustic analysis of the data of twelve speakers. Although a number of studies exist on OCP violations in Bantu languages (see references below), there are only a few acoustic studies available on downstep (see Gjersøe 2015 and contributions in Downing & Rialland 2017; Liberman et al. 1993 is an experimental study on downstep in Igbo, a Kwa language of West Africa).

The article is structured as follows: Section 2 introduces the OCP in Bantu languages and discusses two existing approaches concerning the structural conditions for downstep in the languages of the Sotho-Tswana family. Section 3 presents the elicited-production study that was carried out in order to test which structural condition causes downstep in Tswana. It includes a qualitative analysis of auditory downstep transcription and a quantitative analysis of an acoustic measure of downstep. Section 4 addresses the theoretical account of the data and explores further predictions. Section 5 concludes.

2 The Obligatory Contour Principle

2.1 The OCP in Bantu languages

The Obligatory Contour Principle (OCP; Leben 1973) expresses the dispreference of identical adjacent elements. It was originally proposed for the representation of tones in an autosegmental framework in which tones are organized on separate tiers. For instance, sequences of high-toned syllables must be represented phonologically as in (1a), where a single high tone is multiply associated with a number of adjacent syllables, and not as in (1b), where identical elements, i.e. high tones in case of (1), are adjacent.

(1) *Representation of high tones according to the OCP*

    a. σ σ σ   b. *σ σ σ
           \/ | | |
           H   H H H

The OCP was initially formulated as a morpheme-structure constraint which rules out sequences of identical tones in underlying representations like in (1b). The argument for a preference of a representation like in (1a) comes from Meeussen's rule, which targets a dissimilation process that changes one of two adjacent high tones to low (Odden 1980). In case of (1b), Meeussen's rule predicts a realization of an HLH tone sequence. If, however, a sequence of three high tones is realized, a representation as in (1a) is more likely: If the sequence is associated with only one high tone, Meeussen's rule does not apply as no further (second) high tone is present. The existence of Meeussen's rule in other contexts thus provides a good argument to assume a representation as in (1a). The OCP has since been extended in a variety of ways to also refer to phonological derivations, e.g. by blocking the application of a phonological rule (McCarthy 1986) if the outcome results in adjacency of identical elements or by triggering a phonological rule in order to avoid a surface OCP violation (Yip 1988). The OCP has also been extended from a constraint on morpheme structure to a constraint applying to higher prosodic constituents.

In Bantu tone languages, OCP violations are avoided in a variety of ways, e.g. through deletion of tones, tone movement, blocking of tone spreading, and fusion of two tones into one (see Myers 1997 for an overview). For instance, in the Bantu language Shona, several of these mechanisms are used to avoid OCP violations or 'repair' them (data from Myers 1997).

In some Bantu languages, violations of the OCP lead to downstep, e.g. in Kishambaa, a Bantu language spoken in Tanzania (Odden 1982, 1986). Kishambaa shows a surface contrast between two adjacent high-toned syllables and a sequence of a high tone and a downstepped high tone. The first case has been analysed as the realization of a single high tone associated

with two TBUs (along the representation in (1a)) and the second case as a sequence of two lexical high tones. In the latter case, a downstep occurs between the two lexical high tones, i.e. that the pitch level of the second high tone is lower than that of the first (Odden 1982, 1986).

This contrast is illustrated in (2a) and (2b). Underlining refers to a lexical high tone. Downstep is indicated by the superscript exclamation mark.[1]

(2) *Downstep in Kishambaa* (Odden 1986: 363; Myers 1997: 883 citing Odden 1982)

    a. /ní- kúi/   ní-ˈkúi   'It is a dog.'
                              | |
      COP-dog H_1 H_2

    b. /ní-ki-chí-kóma/   ní-kí-ˈchí-kómá 'I was killing it.'
      1SG-PROG-OC-kill H_1 H_{2+3}

Downstep occurs between two adjacent high tones associated with different morphemes, as in (2a) (although it can also occur in the same morpheme; Odden 1982: 187, 189). The data in (2b) show that no downstep occurs between the surface high tones that are multiply linked to the same underlying tone. Multiple linking in (2b) arises from two tonal processes: First, High Tone Spread from the subject prefix *ní-* onto the following aspect marker *-ki-*; second, high tone fusion of the high tone of the object marker *-chí-* and that of the verb stem *-kóma*. Note that the OCP does not act as a rule blocker for High Tone Spread or tone fusion in this context to avoid the adjacency of H_1 and H_2. High Tone Spread of H_1 takes place although another high-toned syllable follows, namely H_2 on the object prefix *-chí-*. Kishambaa is thus an example of a language that violates the OCP by tolerating adjacent high tones in the surface representation, but resolves their potential non-distinctness by downstepping the second high tone.

To sum up, it is well-documented that the OCP is an important constraint on tonal representations. Many languages obey the OCP at the level of the word or its subconstituents, which leads to tone deletion, blocking of tone spread, and/or fusion. However, in other languages, violations of the OCP lead to a downstepped phonetic implementation.

2.2 The OCP in Sotho-Tswana

Since the current study explores OCP violations across word boundaries in Tswana, the following section provides relevant background to Tswana and a review of OCP effects in the language family Tswana belongs to. Tswana belongs to the Sotho-Tswana group of Southern Bantu languages (S30 in Guthrie's (1967–1971) classification), together with Southern Sotho (Sesotho), Northern Sotho (Sepedi) and Lozi (Silozi). There are several descriptions available in the linguistic literature on aspects of the tonal system of these languages (see Ziervogel,

---

[1] In all examples, surface high tones are indicated by an acute accent, underlying high tones by underlining, and downstep by a superscript exclamation mark. Glossing follows the Leipzig Glossing Rules (http://www.eva.mpg.de/lingua/resources/glossing-rules.php). Numbers in the glosses refer to noun classes. The following abbreviations are used:

AGR	agreement	NEG	negation	PROG	progressive
ASP	aspect	NP	noun prefix	PST	past tense
COP	copula	OC	object concord	REL	relative
DEM	demonstrative	PL	plural	SC	subject concord
LOC	locative	POSS	possessive concord	SG	singular

Lombard & Mokgokong 1969, Lombard 1976, Monareng 1992 for Northern Sotho; Khoali 1991 for Southern Sotho; Mmusi 1992, Chebanne, Creissels & Nkhwa 1997, Creissels 1999, 2000 for Tswana). Tone has a lexical and a grammatical function in Tswana. Like Shona and Kishambaa and most other Bantu languages, Tswana contrasts two level tones, high (H) and low (L) on the surface. As in many other Bantu languages, only high tones are present underlyingly (see Mmusi 1992: 39; Hyman & Monaka 2011 for varieties of Tswana).

Previous research on Tswana has claimed that this language has an active OCP restriction according to which sequences of singly-linked adjacent high tones are allowed but sequences of multiply-linked adjacent high tones are not allowed (Mmusi 1992). Mmusi (1992) cites examples from the domain of the root and word that illustrate the various processes that apply in order to avoid OCP violations. In (3a), the singly-linked lexical high tone of the subject prefix and that of the verb root are both realized.

(3) *OCP in Tswana* (Mmusi 1992: 70, 112)

    a. /ó-réka/ → ó-r$\underline{\acute{e}}$ká nama (SC1-buy meat)   'S/he is buying meat.'
$$\text{ó-r}\underset{\diagdown\!\diagup}{\underline{\acute{e}}}\text{ká}$$
$$H$$

    b. /ó-a-réka/ → ó-a-r$\underline{\acute{e}}$ká (SC1-ASP-buy)   *ó-á-réká   'S/he is buying.'
$$\qquad H_1 \quad H_2 \qquad\qquad H_1 \quad H_2$$

Mmusi (1992) suggests an analysis according to which these separate lexical high tones are fused into one, which is shown in (3a). In (3a), also the rule of High Tone Spread applies. High Tone Spread takes place from a lexical high-toned syllable to an adjacent toneless syllable if this syllable is not phrase-final and if this does not result in adjacent multiply-linked high tones. Example (3b) shows that the OCP acts as a rule blocker for High Tone Spread if it were to create a sequence of multiply-linked high tones.

For the closely related language Southern Sotho, Khoali (1991) also describes tonal processes that counteract violations of the OCP, among them fusion (4a) which merges adjacent lexical high tones within the verb into a multiply-linked high tone (co-occurring with High Tone Spread of the lexical high-toned syllable to the adjacent syllable in (4a)).

(4) *OCP in Southern Sotho varieties* (Khoali 1991: 296)

    a. /ó bóna/ → $\underline{\text{ó}}$-b$\underline{\text{ó}}$ná ... (SC1-see)   'S/he sees ...'   [Lesotho]
$$H$$

    b. /ó-se-kgúrúmetse/ → $\underline{\text{ó}}$-se-kg$\underline{\text{ú}}$r$\underline{\text{u}}$métsé (SC1-NEG-cover) 'S/he does not cover.'
$$\qquad H_1 \qquad H_2\ H_3$$
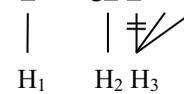
    c. /ó-bóna/ → $\underline{\text{ó}}$-b$\underline{\text{o}}$ná ... (SC1-see)   'S/he sees ...'   [Free State]
$$\qquad H_1\ H_2$$

High Tone Spread and then left-branch delinking occurs with a grammatical high tone (H3) co-occurring with the negated verb form in the verb stem, as in (4b). Khoali (1991) also investigates dialectal variation and observes that some dialects apply left-branch delinking in the verb, as in Southern Sotho spoken in the Free State, (4c), compared to fusion in Southern Sotho spoken in Lesotho, (4a).

2.3 Downstep in Sotho-Tswana

For OCP violations above the word level, downstep has been reported as one resolution strategy, i.e. the phonetically lowered realization of the second in a sequence of two adjacent high tones. There exist two distinct accounts in the literature on the Sotho-Tswana languages which differ in the structural conditions that trigger downstep. They will be introduced in this section.

2.3.1 The tonotactic approach

The 'tonotactic approach' (Creissels 1998; term suggested to us by a reviewer) postulates an exclusively tonally determined context for the occurrence of downstep in Tswana. More precisely, a downstep occurs across word boundaries when a word-final high tone is followed by at least two successive high-toned syllables. 'This downstep occurs irrespective of the precise syntactic nature of the boundary between the two words and of their morphological structure' (Creissels 1998: 146). Examples are provided in (5) below, where, according to Creissels (1998: 145), downstep between the subject and the verb takes place in (5a) and is blocked in the same syntactic environment in (5b) because the tonotactic condition of three adjacent high tones, of which two follow the word boundary, is not met. The final high tone of the first word can be either lexical, or derived by High Tone Spread; Creissels gives examples from the latter (as in (5a, b)).

(5) *Tswana* (Creissels 1998: 145)

    a.   Ba-sádí   ˈbá-bú-íle   tháta.

          NP2-woman   SC2-speak-PST   much

          'The women have spoken a lot.'

    b.   Pódí   é-sulé.   (*Pódí ˈé-sulé.)

          NP9.goat   SC9-die.PST

          'The goat has died.'

2.3.2 The phrasal approach

In contrast, Khoali (1991) postulates a 'phrasal approach' (our term) to downstep in Southern Sotho (see also Kunene 1972 for a descriptive pilot study on downstep in Southern Sotho). Within the framework of Prosodic Domains (Nespor & Vogel 1986), Khoali (1991) states that in a sequence of two adjacent high tones across word boundaries, a word-initial high tone is downstepped if both words occur within the same phonological phrase. Khoali (1991) adopts the indirect reference approach (e.g. Nespor & Vogel 1986; Selkirk 1986, 2011; Truckenbrodt 1995) which proposes that phonology is not directly conditioned by syntactic information. Rather, the interface is mediated by phrasal prosodic constituents, like phonological phrase (abbreviated as φ), which are defined with reference to syntactic constituents but need not match them. Khoali (1991) provides a phrasing algorithm that derives the phonological phrasing from the syntax.[2]

---

[2] 'The P-domain includes all constituents on the non-recursive side of a lexical head as well as all constituents C-commanded by such a head up to the right edge of the last constituent. The head c-commanded by any constituent on its recursive side is at the right edge of the P-domain and the first left constituent of the constituent that C-commands a head on the recursive side is the beginning of the next P-domain. Phonetic modifications in certain syntactic contexts lead one to the belief that Doke's "qualificatives" [i.e. modifiers, SZ] C-command the head noun.' Khoali (1991: 17)

According to Khoali's phrasing algorithm, the subject and a following verb constitute one phonological phrase in Southern Sotho. An example from Southern Sotho with the phrase structure made explicit is shown in (6).

(6) *Southern Sotho* (Khoali 1991: 65, 76)

    a. (((Pódí)$_\omega$ ($^|$é-rátá)$_\omega$ (moríti)$_\omega$)$_\varphi$)$_{IP}$
        NP9. goat SC9-likes   shade
        'A goat likes shade.'

    b. (((Pódí)$_\omega$ ($^|$é-bátla)$_\omega$ (jwang)$_\omega$)$_\varphi$)$_{IP}$
        NP9.goat SC9-want   grass
        'The goat wants grass.'

Khoali's (1991) approach predicts downstep on the high tone of the subject concord in (6a) and also in (6b).

Khoali adduces language-internal and external evidence for the phrasing in (6): Nouns with an underlying lexical high tone on the stem-initial syllable (*pó-* in (6a) and (6b)) are realized with an HL-pattern (as opposed to an HH-pattern) when they occur final in a phonological phrase. Thus, the High Tone Spread to the final syllable of the subject in (6) marks it as not phrase-final (Khoali 1991: 70).

In (7), the noun and following possessive concord are separated by a phonological phrase boundary according to Khoali (1991).

(7) *Absence of downstep in Southern Sotho* (Khoali 1991: 114, 117)

    a. (Di-pódi)$_\varphi$ (tsé-$^|$kgóló   hahóló $^|$dí-fihlile)$_\varphi$
        NP10-goat POSS10-big very   SC10-arrived
        'Very big goats have arrived.'   (→ *Dipódí $^|$tsé ...)

    b. (Bathépu)$_\varphi$ (basélé $^|$bá-fihlile)$_\varphi$
        NP2-Thembu POSS2.different SC2-arrived
        'Different Thembus have arrived.'   (→ *Bathépú basélé ...)

Again, such a phrasing is supported by language-internal evidence. In non-phrase-final position, High Tone Spread to the word-final syllable would have taken place. As shown in (7), the final syllables of the head nouns are low toned, which Khoali interprets as an indication of their phrase-final status.

Interestingly, both approaches predict downstep for the data in (6), yet for different reasons. While for Khoali (1991) the prosodic structure of two words within the same phonological phrase is decisive for downstep, for Creissels (1998) it is the phonological condition of three adjacent high tones. Thus, we have two competing accounts of downstep for the Sotho-Tswana languages by Creissels (1998) and Khoali (1991), and the present study investigates experimentally which of the two accounts holds for the variety of Tswana under investigation. The data of the previous studies are based on elicitation (Creissels 1998) and introspection (Khoali 1991), and they come from different, though closely related languages. It is not clear whether the reported differences are linked to the languages investigated or if they form part of dialectal variation. After all, there is evidence that Tswana shares the phrasing of subject and verb (as in (6) above) and noun and possessive (as in (7a) above) with Southern Sotho, as will be shown and discussed in Section 4.1. Furthermore, both Creissels and Khoali explicitly state that they expect dialectal variation across the different varieties of Sotho-Tswana, including the tone system. The data in Creissels (1998) come from a single speaker

who is a native of the town Kanye and identifies herself as a speaker of the Sengwaketse dialect. In Cole's (1955: xvi) classification, Sengwaketse belongs to the Central division of the Setswana cluster of dialects. Khoali's work investigates the variety of Southern Sotho spoken in Qacha's Nek in Lesotho and the Maluti area of South Africa.

The current study tests both approaches empirically with speakers of Tswana spoken around Vryburg in the North-West province in South Africa. Vryburg is reported to host dialects of the Southern division of Tswana (Cole 1955: xvi). Geographically the three varieties of Sotho-Tswana discussed in this article are thus distributed across an extended area in Southern Africa. By running a production study obtaining acoustic data from several speakers, the current study thus presents the first instrumental approach to the occurrence of downstep in Tswana (as spoken in Vryburg).

2.4 Predictions and hypotheses

In order to test whether the 'tonotactic' or the 'phrasal' approach accounts for downstep patterns in Tswana, we constructed experimental stimuli that systematically take the predictions of the tonotactic and phrasal approaches into account. In particular, two syntactic structures which differ in phonological phrasing, namely subject–verb sequences and noun–possessive sequences are tested with two different tonal contexts. We follow a 2×2 design with tonal context (TONE) and syntactic structure/phonological phrasing (PHRASING) as the two factors, resulting in four contexts as shown in (8) below. As a structural prerequisite, the tonotactic approach requires three adjacent high tones on the surface that are divided such that a word boundary occurs between the first and the second high tone. The phrasal approach, on the other hand, requires at least two adjacent high tones and a word boundary between them. Crucially, for downstep to occur, the two words must belong to the same phonological phrase.

(8)	SYNTACTIC STRUCTURE	PHONOLOGICAL STRUCTURE	TONOTACTIC APPROACH PREDICTION	PHRASAL APPROACH PREDICTION
a.	N SC-V...	$((... \acute{\sigma}_{H1})_\omega\ (\acute{\sigma}_{H2}\ \acute{\sigma}_{H3}\ ...)_\omega...)_\varphi$	ó ꜜó ó ...	ó ꜜó ó ...
b.	N SC-ASP-V	$((... \acute{\sigma}_{H1})_\omega\ (\acute{\sigma}_{H2}\ \sigma\ \acute{\sigma}\ ...)_\omega)_\varphi$	ó ó σ ...	ó ꜜó σ ...
c.	N POSS N SC-V...	$((... \acute{\sigma}_{H1})_\omega)_\varphi\ ((\acute{\sigma}_{H2}\ \acute{\sigma}_{H3}\ ...)_\omega...)_\varphi$	ó ꜜó ó ...	ó ↑ó ó ...
d.	N POSS N SC-V...	$((... \acute{\sigma}_{H1})_\omega)_\varphi\ ((\acute{\sigma}_{H2}\ \sigma\ ...)_\omega...)_\varphi$	ó ↑ó σ ...	ó ó σ ...

The two approaches converge in their predictions for contexts (8a) and (8d). In (8a), a subject noun (N) with a final high tone (H_1) is followed by a lexically high-toned subject concord (SC) (H_2) which itself is followed by a lexically high-toned verb (V) (H_3), as in (5a) above. Both approaches predict downstep to occur before the subject concord, though for different reasons. In (8d), a head noun (N) with a final high tone (H_1) is followed by a high-toned possessive concord marker (POSS) (H_2) which itself is followed by a low-toned noun (N). Both approaches predict downstep not to occur before the possessive concord, again for different reasons.

The two approaches make different predictions for contexts (8b) and (8c) since these differ in their structural conditions, i.e. in tonal structure and phonological phrasing. In (8b), a subject noun (N) with a final high tone (H_1) is followed by a lexical high-toned subject concord (SC) (H_2) which itself is followed by a low-toned aspect marker (ASP). In (8c), three adjacent high tones occur across a word boundary in a possessive construction with a phonological phrase break between the noun and its possessor.

The tonotactic approach (Creissels 1998) predicts downstep to occur in (8c) since three adjacent high tones occur across a word boundary but not in (8b) and (8d) since only two high tones are adjacent across a word boundary. It thus predicts contexts (8a) and (8c) to pattern together (downstep) vis-à-vis contexts (8b) and (8d) (no downstep).

The phrasal approach (Khoali 1991) predicts downstep in contexts (8a) and (8b) due to their parallel phrasal structure. Contexts (8c) and (8d), in contrast, show an intervening phrase boundary induced to the left of a nominal modifier (POSS N) that blocks the application of downstep. It thus predicts contexts (8a) and (8b) to pattern together (downstep) vis-à-vis contexts (8c) and (8d) (no downstep).

3. Production study

3.1 Method

3.1.1 Target sentences

In order to test the predictions outlined in (8), the production experiment followed a 2×2 design, such that two surface tonal conditions appeared in two boundary conditions. The tonal conditions had two adjacent high tones across word boundaries and differed in the third tone, either (i) H)$_\omega$ $_\omega$(H**H**, or (ii) H)$_\omega$ $_\omega$(H**L**. The boundary conditions at the ω-boundary distinguished between the absence and presence of a φ-phrase boundary: H)$_\omega$, and H)$_\omega$)$_\varphi$, respectively.

For each of the contexts A, B, C, and D (corresponding to (8a–d), four different sentences were constructed. Examples are given in (9). A full list can be found in the appendix.[3] The total set of test sentences comprised three repetitions of those 16 sentences.

(9) *Illustration of contexts A, B, C, D*

    a. Ba-rwá   bá-thúsá   ba-lemi.

       NP2-son   SC2-help   NP2-farmer

       'The sons help the farmers.'

    b. Ba-n̲ná   bá-a-húma.

       NP2-man   SC2-ASP-become.rich

       'The men are becoming rich.'

    c. Di-pitsé   tsá-gá-Mó-thibi   dí-pêdí.

       NP10-horse   POSS10-LOC-NP1-NAME   SC10-two

       'Mothibi has two horses.' (lit.: 'The horses of Mothibi are two.')

    d. Di-tsêb̲ế   tsá-kgômó   dí-di-kgólo.

       NP10-ear   POSS10-cow   SC10-AGR10-big

       'The ears of the cow are big.'

---

[3] Here and in the Tswana data in the appendix the following needs to be noted: Nasals can be syllabic in Sotho-Tswana and carry tone, as in (9b), for example. Circumflex on <o> and <e>, as in examples (9c, d), indicate a difference in vowel quality concerning differences in vowel height, i.e. vowels with a circumflex are lower than the ones without.

Although considerable care was taken in the construction of the stimuli, a number of asymmetries emerged at the locus of interest (final syllable of the noun (H1) and first syllable of the following word (H2)). These are listed here: (i) The stimuli vary in the number of syllables that precede the locus of interest in contexts A, B and C between one and three. (ii) In contexts A and B, a voiced obstruent intervenes in the locus of interest, whereas it is a voiceless obstruent in contexts C and D. (iii) There are lexical high tones and derived high tones in the locus of interest (on the verb in context A, on the noun in context B). Remember that this should not make a difference; see example (6) above, for instance, for a case of a derived high tone that causes downstep. (iv) The number of low-toned syllables following H2 in context D varies between one and two, and shows even a syllabic nasal in one case. (v) In contexts B, C and D, there are obstruents before H1, whereas it is sonorants in context A. Despite these asymmetries, we believe that none of them affect the presence or absence of downstep. Some of these factors may affect the acoustic analysis of downstep since microprosodic influences due to different types of segments show up in f0. However, this was taken care of by applying different types of f0 analyses (see Section 3.3.1 below).

3.1.2 Recording procedure and participants

The 48 sentences (4 contexts $\times$ 4 sentences $\times$ 3 repetitions) were presented in pseudo-randomized order using presentation software, interspersed with a small number of fillers showing a similar syntactic structure. Stimuli were presented in Tswana only, using Tswana orthography which does not mark tone. Care was taken that the same contexts did not occur twice after each other. Participants were given the possibility to read through the sentences first to familiarize themselves with the sentences. They were then asked to read out the sentences from the screen one-by-one. There was a short break after one third of the full set of stimuli.

Recordings were done in people's homes in Vryburg and Huhudi, North-West Province, South Africa. An M-audio microtrack II recording device was used (sampling rate 44.1 kHz) together with a head-mounted microphone. Care was taken to minimize background noise. Data were transferred to a computer hard drive for further analysis.

Twelve speakers (five male and seven female, aged between 15 years and about 50 years) took part in the study. Speakers whose speech was considered representative for the area by the local pastor were invited by him to take part in the research voluntarily. According to Cole (1955: xvi), the variety of Tswana spoken in this region belongs to the Southern division of Tswana. In total, 576 sentences were recorded (4 contexts $\times$ 4 sentences $\times$ 3 repetitions $\times$ 12 speakers).

3.2 Qualitative analysis

Of the data collected, 530 sentences were annotated based on auditory impression (46 sentences had to be discarded due to disfluencies or mispronunciations). The question asked during annotation was if from the two high-toned syllables under consideration, the second syllable is perceived as downstepped. Both authors of the study listened to the sound files of all speakers in all conditions and transcribed each token according to this question. The transcription was first done independently of each other and was based solely on auditory impression. Information on actual fundamental frequency (f0), e.g. by means of a pitch track, was not taken into consideration. Both authors are phonetically-trained and have experience with African tone languages.

Each target sentence thus received two judgments concerning the occurrence of downstep, one from each labeller. The inter-annotator reliability, i.e. the agreement between the two judgments across the whole data set, was calculated after this initial round of annotation

Table 1 Count of number of disagreements per speaker.

Speaker	01	02	03	04	05	06	07	08	09	10	11	12	Total
Count	6	4	5	11	11	15	5	1	7	6	5	6	82

Table 2 Auditory impression per speaker and context, giving the number of downstep/no downstep/unresolved disagreement.

Speaker	Context A	Context B	Context C	Context D
01	10/1/1	12/0/0	0/12/0	0/10/0
02	8/1/1	11/0/0	1/8/0	0/5/0
03	7/3/2	12/0/0	3/9/0	3/7/0
04	5/7/0	10/2/0	0/11/1	0/12/0
05	0/12/0	8/4/0	1/10/0	0/11/0
06	5/5/1	11/1/0	2/10/0	0/11/0
07	0/9/1	12/0/0	0/9/0	0/8/0
08	12/0/0	11/1/0	0/12/0	0/12/0
09	2/8/1	11/0/0	0/11/0	0/9/0
10	7/3/0	12/0/0	0/10/0	0/9/0
11	8/3/1	12/0/0	0/11/0	0/11/0
12	7/2/2	9/0/1	0/9/1	0/12/0

using Cohen's kappa (weights: unweighted). The resulting $\kappa = 0.683$ ($z = 16.1$, $p = 0$) shows substantial agreement (see Landis & Koch 1977 for the interpretation of kappa). The total agreement was 84.3%.

There were 82 cases of disagreement in the first round of annotation (i.e. 15.7%). As we had a closer look into their distribution, we will briefly comment on them here. Table 1 shows the disagreements per speakers.

The number of disagreements also differed across the different contexts (A = 30, B = 25, C = 11, D = 16). Presumably, segmental characteristics might have contributed to the disagreement. In 36 cases, disagreement occurred when the vowels of the two high tones in question differed in vowel quality. In this case, the vowel of H1 was always a front high or mid vowel (/i/ or /e/) and the vowel of H2 was the low, back vowel /a/, as in the sentences A4, B3, B4, C1, and D4 (see Appendix). Vowel-intrinsic pitch is lower in low vowels (e.g. Whalen & Levitt 1995). In 14 cases, the vowel in H2 was preceded by the affricate /ts/ which was realized as an ejective, leading to clearly perceivable coarticulated laryngealization on the vowel. We will not pursue this further though, given that speaker-specific and segmental influences on tone perception lie outside the scope of the present paper and agreement could be reached in nearly all cases after discussion.

Table 2 lists the reported occurrence of downstep, no downstep, and unresolved disagreements between the two annotators in the different contexts per speaker.

There is a clear pattern that emerges across all speakers, namely that downstep occurs in context B but not in context C or context D. In context A, we find speaker-specific variation since a clear pattern of downstep is not perceived in all speakers. Speakers 04, 05, 06, 07, and 09 are not perceived to consistently produce downstep. We take up the speaker-specific variation in Section 3.3.2 below. The quantitative analysis will be discussed in the next section.

3.3 Quantitative analysis

3.3.1 Data analysis

The main acoustic correlate of tone is fundamental frequency (f0) (see Myers 1998), which can only be realized on voiced segments. In order to guarantee comparability across all stimulus items, the vowels of the high-toned syllables were delineated and analysed for their f0, following segmentation criteria in Turk, Nakai & Sugahara (2006), except for double vowels which we divided by half of their total duration. For every sentence, the vowels carrying the noun-final high tone and the following word-initial high tone were the target vowels used for analysis.

Studies on the acoustic realization of a single high tone in the closely related Bantu language Northern Sotho have shown that fundamental frequency starts rising on the tone-bearing syllable but the f0 peak is only reached later (Zerbian & Barnard 2009). Its exact location varies according to a number of factors: It is reached earlier in syllables containing a voiceless onset than in syllables with a sonorant onset, it occurs earlier in utterance-initial onsetless syllables, and its exact location is speaker-dependent. Frequently, however, the f0 peak is only found between two or three syllables away, counting from the tone-bearing syllable. This behaviour has been described as High Tone Spread in the literature, and occurs when no low tone target follows the high-toned syllable.

Peak spread into the adjacent syllable as in High Tone Spread is not expected to occur in the present data as the second syllable is associated with a lexical high tone and has thus its own tonal target. We thus expect the tonal target to be realized locally.

It remains an open question where to measure the acoustic correlate of the tone of a syllable in such a case. In the following we motivate our choice. Myers (1998, 1999) used local f0 maxima in his study on the alignment and scaling of high tones in Chichewa. Zerbian & Barnard (2009, 2010) displayed time-normalized pitch contours across relevant syllables in their study on Northern Sotho and used the average f0 of a vowel for the statistical analysis. For the exposition of our acoustic results, we compared the three measures in (10), based on the respective vowels' f0 average.

(10) *Measures for comparing the tone of the final vowel of the noun (H1) to the initial vowel of the following word (H2)*
   a. Mean f0 of H1 to mean f0 of H2
   b. Mean f0 of central 50% of H1 to mean f0 of central 50% of H2
   c. Mean f0 of final third of H1 to mean f0 of final third of H2

Both the local f0 maximum (Myers 1998) and the mean f0 over a vowel (measure (10a), Zerbian & Barnard 2009, 2010) are susceptible to microprosodic disturbances induced by segments and/or tones (Lehiste & Peterson 1961, Hombert, Ohala & Ewan 1979). While the latter is controlled for in the study and at least kept constant within contexts, the former could not be fully controlled due to restrictions posed by the language itself. For instance, there are few subject concords in Tswana which start with a sonorant. It was therefore unavoidable that we use subject concords starting with a voiced plosive which is known to lower the f0 slightly (Hombert et al. 1979). Similarly, in the possessive construction some concords start with voiceless plosives which raise the f0 slightly (Lehiste & Peterson 1961).

In measure (10b), the microprosodic influences by preceding and/or following segments and tones are cut off by using the mean f0 of only the central 50% of the vowel. However, it is known since Myers' (1999) study that the pitch peak of a high-toned syllable in Bantu can be reached late in the syllable or even only in the following syllable. Zerbian & Barnard (2009) showed that this holds if the syllable carrying the high tone has a sonorant onset. Thus, using only the central 50% for analysis might leave out the actual pitch target from

Table 3 Comparison of three phonetic measures for phonological tone in Hz.

Context	Overall f0 mean		f0 mean last third		f0 mean mid-50%	
	H1	H2	H1	H2	H1	H2
A	185.0	176.5	186.6	171.7	185.6	176.1
B	181.5	164.3	179.9	160.6	181.9	163.7
C	174.5	176.0	171.7	174.3	174.2	176.0
D	178.5	182.0	175.9	182.2	178.2	182.0
Context	Std.dev. mean		Std.dev. last third		Std.dev. mid-50%	
	H1	H2	H1	H2	H1	H2
A	53.9	50.2	54.1	48.9	54.1	50.0
B	52.6	45.4	52.4	45.6	52.6	45.0
C	47.7	47.5	46.4	47.4	47.6	47.2
D	48.6	47.8	47.0	49.0	48.2	47.7

analysis. This would speak in favour of measure (10c), which considers the f0 only in the last third of the vowel, thereby cutting off microprosodic influences of a preceding segment but retaining a possible pitch target towards the end of the syllable. However, the segment boundary and thus duration of the vowels in context B (*bá-a-* . . .) could not be determined reliably. Additionally, the long vowel was often accompanied with creaky voice (and hence undefined f0 values) halfway through the vowel. These facts would not affect measure (10b), which looks at the mean f0 of the central 50% around the mid of the vowel.

For the quantitative analysis items were excluded which contained hesitations, mis-pronunciations, or pauses affecting the syllables under consideration ($n = 59$), and/or a considerable amount of creaky voice ($n = 13$) which does not allow f0 measures. Hence, the quantitative analysis was based on 517 cases. A comparison of the means and standard deviations of the three measures in (10) shows that the differences between the measures are not substantial (see Table 3).

Given that the differences between the measures are not substantial, the mean of the mid-50% of the tone-bearing vowel is used as acoustic reference because it cuts off microprosodic influences. The further analysis will be based on this measure.

As a reviewer suggested, an alternative method of measuring f0 contours is to fit them to polynomial curves based on f0 values at landmarks in the contour, such as onset of rise, peak, plateau, or offset of fall (e.g. Andruski & Castello 2004). Given that Tswana is a tone language with only two level tones and little tonal density, the mentioned landmarks are not necessarily local or salient. Figures 1 and 3 below illustrate the problem of defining suitable landmarks for tones in the current study. Whereas f0 maximum seems an appropriate landmark of the tonal target for H1 in context A, in context C, the f0 maximum of H1 clearly shows undesirable but unavoidable microprosodic influences of the affricate. We thus decided against such a modelling approach also because the main focus of the current study is the occurrence of phonological downstep, not the detailed phonetic modelling of the contour. Moreover, our acoustic measures are to be seen in conjunction with the auditory transcription reported in the previous section.

The presence or absence of downstep is established through a comparison between the f0 of H1 and H2 (see Snider 1998, 2007 for downstep in Bimoba and Chumburung, and Dolphyne 1994, Genzel & Kügler 2011 for downstep in Akan). Downstep occurs when H2 is systematically lower than H1.

Table 4 Estimation of downstep realization between H1 and H2 for each speaker in context A.

Speaker	Gender	Age	H1 (Hz)	H2 (Hz)	H1–H2 difference	Statistics
01	F	~50	228	194	***	$t = 7.8$, df $= 11$, $p < .001$
02	M	~20	99	94	**	$t = 4.1$, df $= 9$, $p < .01$
03	F	~20	208	199	***	$t = 6.3$, df $= 11$, $p < .001$
04	M	~30	145	143	n.s.	$t = 1.6$, df $= 11$, $p = .1371$
05	M	~17	136	141	**	$t = -3.8$, df $= 11$, $p < .01$
06	F	~15	242	233	n.s.	$t = 2.2$, df $= 10$, $p = .05337$
07	M	~45	126	130	**	$t = -3.9$, df $= 9$, $p < .01$
08	M	~35	126	110	***	$t = 8.5$, df $= 11$, $p < .001$
09	F	~20	225	221	n.s.	$t = 1.2$, df $= 10$, $p = .2740$
10	F	~50	258	242	***	$t = 6.1$, df $= 9$, $p < .001$
11	F	~20	244	230	*	$t = 2.4$, df $= 11$, $p < .05$
12	F	~50	178	169	*	$t = 3.0$, df $= 10$, $p < .05$

Significance levels: *.05, **.01, ***.001

3.3.2 Inter-speaker variability

In context A, both the tonotactic and the phrasal approach predict downstep between the final high tone of the first word (H1) and the first high tone of the second word (H2). We thus expected to find the comparison between H1 and H2 to be significantly different across all speakers, which means that downstep is realized. Surprisingly, we found considerable inter-speaker variability in this context (only), which we report on in this section. A series of paired samples *t*-tests were run examining for each speaker whether H1 and H2 were significantly downstepped in context A. Seven out of 12 speakers realized a significantly downstepped H2 (see Table 4). One speaker (speaker 06) very closely approached the significance level of $p < .05$. The results from auditory impression mirror the statistical results: There is an equal number of perceived downsteps and non-downsteps in context A, furthermore there was a high number of disagreements in this context. For two speakers (speakers 04 and 09) the difference between H1 and H2 in context A is not significant, which we conclude to be an indication that downstep was not realized. This is confirmed by the auditory impression as these speakers realize more contours that were perceived as non-downstepped than as downstepped. Two other speakers (speakers 05 and 07) show a significant effect, though in the other direction (indicated by negative *t*-values). Again, this result is mirrored in the results from auditory impression because these two speakers are not perceived as realizing downstep at all.

The speakers who did not produce downstep in context A pattern with the other speakers in contexts B, C, and D though as will be evident later. There is no comparable inter-speaker variation in contexts B, C, and D. This means that the speakers do produce downstep in context B consistently, as well as they did produce no downstep in contexts C and D. The reason for the high inter-speaker variability in context A remains unclear, neither age nor gender seems to be decisive.

3.3.3 Results

This section presents the results of the analysis of the mean f0 in the central 50% of the high-toned vowels H1 and H2. It starts out with representative illustrations, followed by descriptive statistics before it turns to testing the hypotheses inferentially. Figures 1–4 exemplify the realization of each context. In each figure, the oscillogram, the spectrogram, the pitch track, orthographic words and target vowels are shown.

The pitch track in Figure 1 shows that the lexical high tone H1 is realized at a higher pitch (mean of mid-50%: 235 Hz) than the lexical high tone H2 (179 Hz).
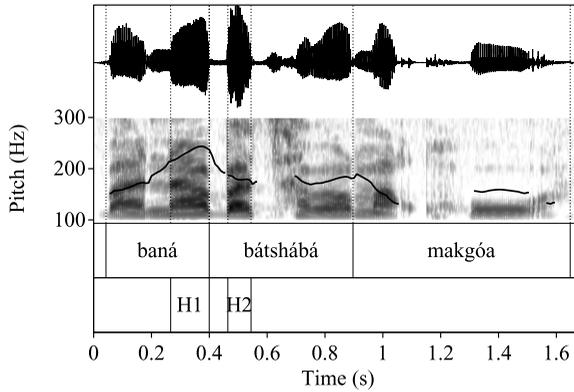
Figure 1 Context A, sentence 2 from speaker O1 ('The children fear the White people').
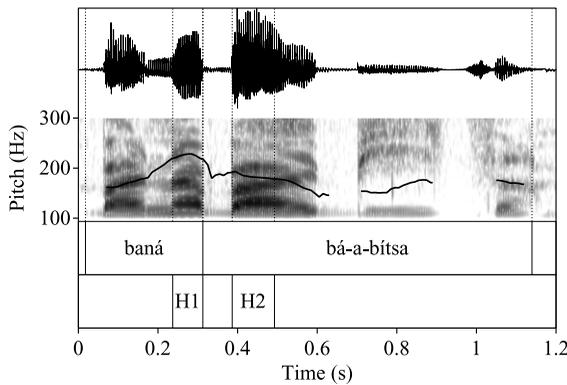
Figure 2 Context B, sentence 2 from speaker O1 ('The children are calling').
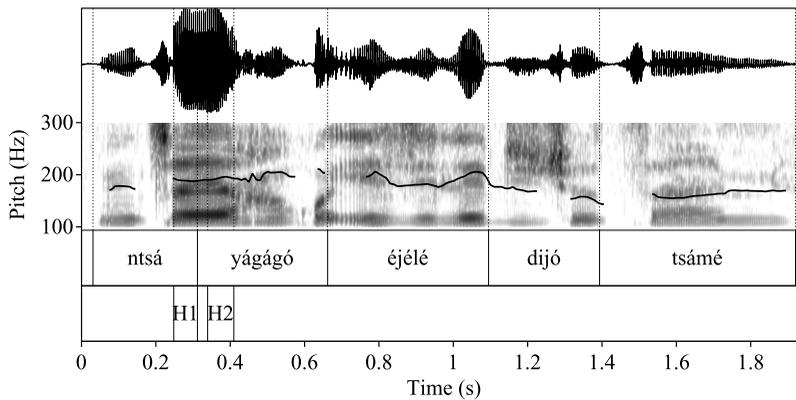
Figure 3 Context C, sentence 3 from speaker O1 ('Your dog has eaten my food').

The example in Figure 2 represents context B. Similar to Figure 1, the lexical high tone H1 is realized at a higher pitch (mean of mid-50%: 227 Hz) than the lexical high tone H2 (183 Hz), which in turn is realized higher than the following low-toned *a*.

The example in Figure 3 represents context C. We see in the pitch track that both lexical high tones H1 (mean of mid-50%: 189 Hz) and H2 (194 Hz) are realized at roughly equal pitch heights.
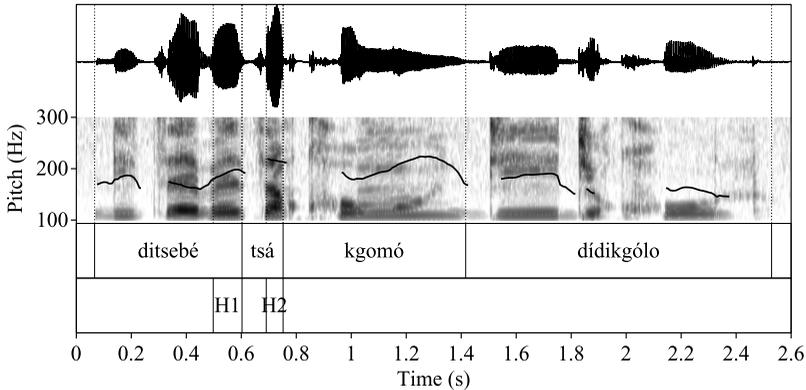
Figure 4 Context D, sentence 1 from speaker O1 ('The ears of the cow are large').
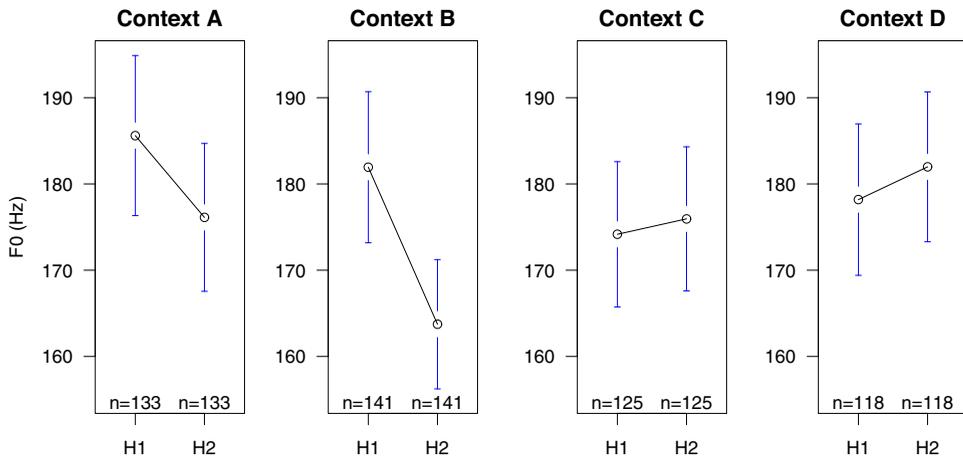
Figure 5 Averaged f0 means measured in the central 50% of the vowels on H1 and H2 across all speakers split by contexts.

The example in Figure 4 represents context D. Tone H2 (mean of mid-50%: 215 Hz) is clearly realized with a higher pitch than H1 (193 Hz) at a high level.

Turning to descriptive statistics, Figure 5 shows the mean values of H1 and H2 in the mid-50% of the vowels respectively in the four contexts, averaged across all twelve speakers (comprising male and female). The black circles give the average mean f0, the blue lines the 95% confidence interval.

Figure 5 shows that, on average, there is a considerable drop in f0 from H1 to H2 in context A and B whereas f0 stays level or rises slightly in context C and D. The exact numerical values for the means together with the standard deviations are given in Table 3 above, right-most two columns.[4] Figure 5 suggests that downstep takes place in contexts A and B but not in C and D.

---

[4] The question was raised whether the downstep is a depressor effect due to the voiced plosive [b] which precedes all H2s in context A and B (but not in context C). Voiced plosives are known to lower the f0 of a following sonorant (House & Fairbanks 1953). However, this microprosodic effect is in the region of at most 7 Hz, thus considerably smaller than the difference of on average 14–21 Hz observed in our data. In the Southern African languages of the Nguni and Shona group, the pitch lowering effect of voiced obstruents has developed into clear depressor effects on the tone in the region of 40 Hz (Traill, Khumalo & Fridjhon 1987). No depressor consonants have been reported for Sotho-Tswana.

Table 5 Report of the linear mixed-effects model specified in the text with the f0 difference between H1 and H2 in Hz as dependent variable.

	Coefficients	SE	*t*-value	Significant at $p < .05$
(Intercept)	9.0396	3.9692	2.277	∗
Phrasing = two phrases	−16.6214	3.0219	−5.500	∗
Tones = two tones	3.6830	1.5641	2.355	∗
Interaction	5.4760	0.7444	7.356	∗
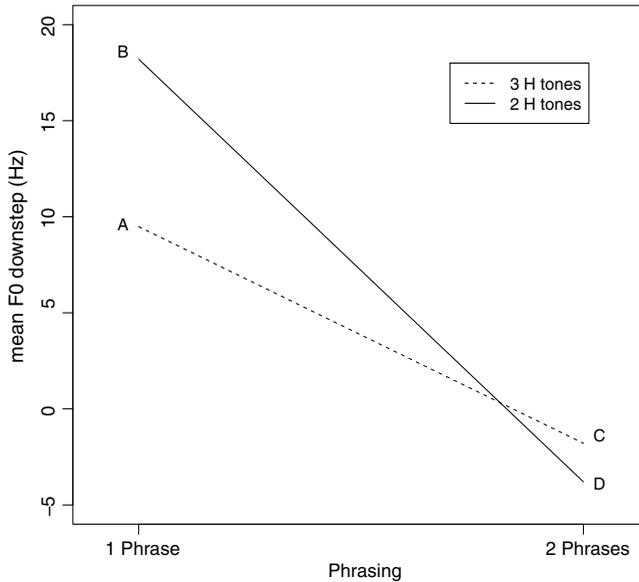
Figure 6 Interaction plot of fixed factors PHRASING and TONES, illustrating the mean f0 difference between H1 and H2 in Hz as an indicator of downstep as dependent variable. The negative amount of downstep in the two-phrases condition illustrates no downstep with a slight increase of f0 between the two high tones.

For the statistical analysis, we fit a multilevel model (Bates et al. 2015) using crossed random factors SPEAKER and ITEM applying random intercepts and slopes for SPEAKER. PHRASING (with two levels 'one phrase'/'two phrases') and TONES (with two levels 'two high tones'/'three high tones') were used as fixed factors. The analysis relied on the difference in f0 between H1 and H2 in Hz as a dependent variable. Contrast-coding was applied with level 'one phrase' as baseline of the factor PHRASING, and level 'three high tones' as baseline of the factor TONES. Model comparisons for the random effect structure was applied. Backward modelling (Barr et al. 2013) of random slopes for SPEAKER was applied, and likelihood ratio tests were run to evaluate the models. The basis for removing factors was set at a *p*-value of the likelihood ratio test of $p < .05$ and lower AIC values. However, the maximal model turned out to be the best fit model. This model assumes that differences exist for each speaker's individual realization of the items.

As shown in Table 5, we find a significant effect for the factors PHRASING and TONES as well as a significant interaction. From the interaction plot in Figure 6 it becomes clear that the factor PHRASING is significant because both contexts A and B differ from contexts C and D in that they show a difference in f0 between H1 and H2 that amounts to 18 Hz on average indicating that the two high tones are downstepped. The mean f0 difference between H1 and H2 in Hz, i.e. downstep, in context A is smaller than in context B due to two facts: First, the

following tone differs, which is H in context A and L in context B. Second, in context A only eight of the twelve speakers realized a difference in f0 between H1 and H2. Four speakers realize no difference of f0 between the two vowels which results in an overall lower mean difference. Contexts C and D show an average f0 difference between H1 and H2 of -1.8 to -3.8 Hz indicating that the two adjacent high tones are not downstepped. The fact that the factor TONES becomes significant is presumably due to the large difference between contexts A and B. Relevant for interpretation is the significant interaction (see Figure 6), which clearly shows that downstep is realized in contexts A and B, and not in contexts C and D.

4 Discussion

The study presented is one of the few that addresses OCP violations above the word level, and furthermore to our knowledge the first study to provide acoustic data on downstep in a Bantu language. The acoustic data confirm the results reached at by auditory impression. In addition to the categorical difference, they reveal a systematic difference in the magnitude of downstep depending on tonal context. We interpret downstep thus as a repair strategy to OCP violations above the word level in Tswana, similar to Kishambaa, a Bantu language spoken in Tanzania (Odden 1982, 1986), and the Zambian Bantu language Namwanga (Bickmore 2000: 302).

Our qualitative and quantitative analyses thus confirm for a Southern division variety of Tswana what has been claimed for closely related Southern Sotho previously, namely that in the realization of two adjacent high tones across word boundaries the occurrence of downstep depends on the syntactic structure, and in turn on the prosodic structure. More precisely, downstep takes place between adjacent high tones belonging to a subject and a verb, which are mapped into one phonological phrase, but not between high tones belonging to a noun and a following modifier, which are mapped into two distinct phonological phrases. Our data thus provide empirical evidence that from the two competing approaches to downstep in the Sotho-Tswana languages which we presented in Section 2.3 the phrasal approach makes the correct predictions for the variety of Tswana under investigation.

This section first discusses the phrase-based analysis put forward by Khoali (1991) for downstep in Southern Sotho before it discusses predictions for an additional syntactic context.

4.1 Phrase-based analysis for downstep

The four contexts investigated in the current study differ in their syntactic structure, and hence in the phonological constituency derived from it (Nespor & Vogel 1986): Following Khoali (1991), in this language, subject and verb form one phonological phrase whereas head noun and modifier form two separate phonological phrases (for the concrete phrasing algorithm see footnote 2). Downstep can then be reformulated as a phonological rule which takes place if the relevant tonal context is met within a phonological phrase (e.g. between subject and verb) but crucially not across phonological phrase boundaries (e.g. nominal head and modifier).

The phrasing put forward here in which the subject constitutes one phonological phrase with the verb is not the default pattern cross-linguistically and thus needs further motivation. It finds its parallels in other Bantu languages (e.g. Kanerva 1990 for Chichewa; Jokweni 1995 for Xhosa; Cheng & Downing 2007, 2009 for Zulu; Zerbian 2007 for Northern Sotho). The phrasing is supported by a tonal rule in Tswana parallel to the one exemplified for Southern Sotho in (7): The last syllable of a phonological phrase is exempt as a target for High Tone Spread in Tswana. The examples in (11) from Tswana show that this is the case in a noun

monna 'man' followed by a modifier, (11b), but not if this noun is a subject followed by a verb, (11a).

(11) *Tswana* (Cole & Mokaila 1962: 7, 83)

    a.   (Mo-<u>n</u>ná   <u>ó</u>-dísá   dí-kgômó)_φ.[5]

        NP1-man   SC1-herd   NP10-cow

        'The man herds the cattle.'

    b.   (Mo-<u>n</u>na)_φ (y<u>ó</u>-<u>ó</u>-nê-<u>n</u>g   á-tsámaya   m<u>ó</u>-tsel<u>é</u>-ng)_φ (k<u>é</u>-rr<u>é</u>)_φ.

        NP1-man   DEM1-SC1-PST-REL   SC1-walk   LOC-path-LOC   COP-1.father

        'The man who was walking along the road is my father.'

The postulated intervening phrase boundary between a noun and its modifier is also not the most common pattern but also finds parallels in other languages, e.g. in the Bantu language Bàsàá (A43, spoken in Cameroon; Hamlaoui, Gjersøe & Makasso 2014). It is furthermore supported by tonal evidence in Tswana, see (11b).

    This generalization concerning the occurrence of downstep in Tswana makes the correct predictions for the phonetic realization of adjacent high tones in verb–object sequences as will be shown in the following section.

4.2 Predictions for other syntactic structures

Verb–object sequences constitute a syntactic configuration in which both constituents are phrased together in a phonological phrase in many Bantu languages (e.g. Kanerva 1990 for Chichewa, Jokweni 1995 for Xhosa, Zerbian 2007 for Northern Sotho). The prediction is that, given two adjacent high tones across word boundaries, downstep will occur on the initial syllable of the object because no phrase boundaries intervene. The example in (12) illustrates this prediction.

(12) *Downstep between verb and object in Tswana*

    (Ke   bítsá   [!]pódi   bosígo)_φ

    1SG   call   NP9.goat   evening

    'I call the goat in the evening.'

Four target sentences with adjacent high tones between verb and object, as in (12), were recorded from the same speakers (again in three repetitions) in order to empirically test the prediction concerning the occurrence of downstep. A representative pitch track is given in Figure 7. The example in Figure 7 clearly shows that H2 (mean of mid-50%: 197 Hz) is downstepped in relation to H1 (228 Hz), as predicted by the phrase-based approach.

    A qualitative analysis of this type of sentences (following the procedure detailed in Section 3.2) revealed the following results: Of the overall 144 recorded sentences (12 speakers × 4 target sentences × 3 repetitions), six had to be excluded due to hesitations. For the remaining 138 sentences, both annotators agreed on the occurrence of downstep in 129 cases. In two cases, both annotators perceived no downstep, and in seven cases there was disagreement in the first round of annotations.

    The predictions made by the phrase-based approach concerning the occurrence of downstep in verb–object sequences are thus borne out in the perceptual analysis, and we expect downstep to occur in other syntactic structures too that meet the respective phonological, i.e. tonal and phrasal, requirements.

---

[5] Note that Cole & Mokaila (1962) do not transcribe downstep in their publication.
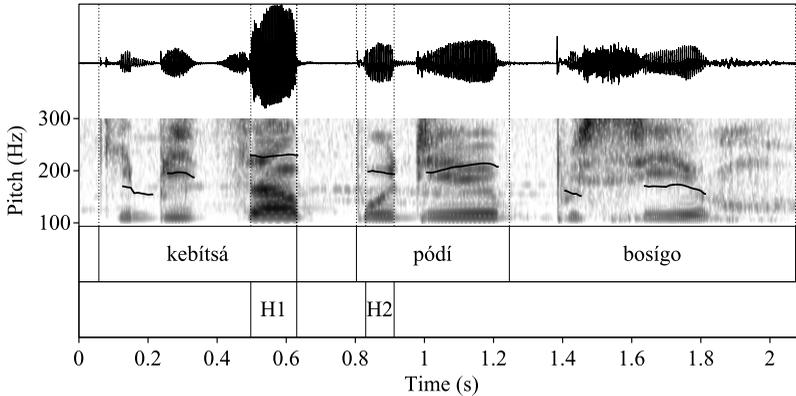
Figure 7 Context of a verb–object sequence, sentence 1 from speaker O1 ('I call the goat in the evening').

5 Conclusion

The article started with the discussion of the Obligatory Contour Principle (OCP; Leben 1973), a cross-linguistically relevant principle that prohibits adjacent identical elements. For Sotho-Tswana there is consensus among Khoali (1991) and Mmusi (1992) that these languages obey the OCP at the word level. The current study has provided the first instrumental data that show how OCP violations are resolved at the phrase-level: Across words within a phonological phrase, OCP violations are resolved by downstep in the Tswana variety investigated in this study, whereas across words and across phonological phrases, two adjacent high tones are realized without downstep.

Acknowledgements

Appendix. Overview of target sentences

Context A (see example (8a) in the text)

1	*Ba-rwá*	*bá-thúsá*	*ba-lemi.*	'The sons help the farmers.'
	NP2-son	SC2-help	NP2-farmer	
2	*Ba-ná*	*bá-tshábá*	*Ma-kgóa.*	'The children fear the White people.'
	NP2-child	SC2-fear	NP6-white.person	
3	*Bó-mmé*	*bá-bóná*	*ba-lé.*	'The mothers see the uninitiated girls.'
	NP2a-mother	SC2-see	NP2-girl	
4	*Ba-símané*	*bá-dísá*	*dí-kgomó.*	'The boys are herding the cattle.'
	NP2-boy	SC2-herd	NP2-cow	

Context B (see example (8b) in the text)

1 *Ba-ńná*   *bá-a-húma.*   'The men are becoming rich.'

   NP2-man   SC2-ASP-become.rich

2 *Ba-ná*   *bá-a-bítsa.*   'The children are calling. '

   NP2-child   SC2-ASP-call

3 *Ba-sádí*   *bá-a-róka.*   'The women are sewing.'

   NP2-woman   SC2-ASP-sew

4 *Ba-rútí*   *bá-a-rútá.*   'The teachers are teaching.'

   NP2-teacher   SC2-ASP-teach

Context C (see example (8c) in the text)

1 *Di-pitsé*   *tsá-gá-Mó-thibi*   *dí-pêdí.*   'The horses of Mothibi are two.'

   NP10-horse   POSS10-LOC-NP1-NAME   SC10-two

2 *Me-raká*   *wá-gá-bó-rré*   *ga-é-kgákála.* 'The cattleposts of my fathers are not far.'

   NP4-cattlepost POSS4-LOC-NP2a-father NEG-SC4-far

3 *Ntšá*   *yá-gágó*   *é-jélé*   *di-jó*   *tsá-mé.* 'Your dog has eaten my food.'

   NP9.dog POSS-2SG SC9-eat.PST NP8-food POSS8-1SG

4 *Di-nkú*   *tsá-kgósí*   *dí-di-ńtsi.*   'The sheep of the chief are many.'

   NP10-sheep POSS10-chief SC10-AGR10-many

Context D (see example (8d) in the text)

1 *Di-tsêbé*   *tsá-kgômó*   *dí-di-kgólo.*   'The ears of the cow are big.'

   NP10-ear   POSS10-cow   SC10-AGR10-big

2 *Di-ntšá*   *tsá-sególê*   *dí-ma-šwé.*   'The dogs of the cripple are ugly.'

   NP10-dog   POSS10-cripple SC10-NP6-ugly

3 *Me-lató* *yá-mmúsó*   *é-me-gólo.*   'The faults of the government are big.'

   NP3-fault POSS10-government SC3-AGR3-big

4 *Mo-lwêtsé* *wá-ngaka*   *ó-mó-gáé.*   'The patient of the doctor is at home.'

   NP1-patient POSS1-doctor   SC1-AGR1-home

References

Andruski, Jean E. & James Costello. 2004. Using polynomial equations to model pitch contour shape in lexical tones. *Journal of the International Phonetic Association* 34, 125–140.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3), 255–278.

Bates, Douglas, Martin Mächler, Benjamin Bolker & Steven Walker. 2015. Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software* 67, 1–48.

Bickmore, Lee S. 2000. Downstep and fusion in Namwanga. *Phonology* 17, 297–331.

Chebanne, Andrew M., Denis Creissels & H. W. Nkhwa. 1997. *Tonal morphology of the Setswana verb*. Munich: Lincom.

Cheng, Lisa & Laura J. Downing. 2007. The phonology and syntax of relative clauses in Zulu. *Bantu in Bloomsbury: Special issue on Bantu linguistics. School of Oriental and African Studies Working Papers in Linguistics* 15, 51–63.

Cheng, Lisa & Laura J. Downing. 2009. Where's the topic in Zulu? In Geertje van Bergen & Helen de Hoop (eds.), *Topics cross-linguistically*: Special issue of *The Linguistic Review* 26(2–3), 207–238.

Cole, Desmond T. 1955. *An introduction to Tswana grammar*. Cape Town: Longman.

Cole, Desmond T. & Dingaan M. Mokaila. 1962. *A course in Tswana*. Washington, D.C.: Georgetown University.

Creissels, Denis. 1998. Expansion and retraction of high tone domains in Setswana. In Larry Hyman & Charles Kisseberth (eds.), *Theoretical aspects of Bantu tone*, 133–194. Stanford, CA: Center for the Study of Language and Information (CSLI).

Creissels, Denis. 1999. The role of tone in the conjugation of Setswana. In Jean A. Blanchon & Denis Creissels (eds.), *Issues in Bantu tonology*, 109–152. Köln: Rüdiger Köppe.

Creissels, Denis. 2000. A domain-based approach to Setswana tone. In Ekkehard Wolff & Orin Gensler (eds.), *Proceedings of the 2nd World Congress of African Linguistics*, 311–321. Köln: Rüdiger Köppe.

Dolphyne, Florence A. 1994. A phonetic and phonological study of downdrift and downstep in Akan. Presented at the 25th Annual Conference on African Linguistics, Rutgers University.

Downing, Laura J. 2011. Bantu tone. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, vol. V, 2730–2753. Oxford: Blackwell.

Downing, Laura J. & Annie Rialland (eds.). 2017. *Intonation in African tone languages*. Berlin: de Gruyter.

Genzel, Susanne & Frank Kügler. 2011. Phonetic realization of automatic (downdrift) and non-automatic downstep in Akan. *Proceedings of the 17th International Congress of Phonetic Sciences* (ICPhS XVII), Hong Kong, 735–738.

Gjersøe, Siri M. 2015. Downstep and phonological phrases in Kikuyu. Masters thesis, Humboldt University Berlin

Goldsmith, John. 1984. Tone and accent in Tonga. In George N. Clements & John Goldsmith (eds.), *Autosegmental studies in Bantu tone*, 19–51. Dordrecht: Foris.

Guthrie, Malcolm. 1967–1971. *Comparative Bantu: An introduction to the comparative linguistics and prehistory of the Bantu languages*, 4 vols. Farnborough: Gregg International.

Hamlaoui, Fatima, Siri M. Gjersøe & Emmanuel-Moselly Makasso. 2014. High tone spreading and phonological phrases in Bàsàá. In C. Gussenhoven, Y. Chen & D. Dediu (eds.), *Proceedings of 4th International Symposium on Tonal Aspects of Languages*, Nijmegen, Netherlands, 27–31. ISCA Archive, http://www.isca-speech.org/archive/tal_2014.

Hombert, Jean-Marie, John J. Ohala & William G. Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55(1), 37–58.

House, Arthur S. & Grant Fairbanks. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America* 25, 105–113.

Hyman, Larry M. 2001. Privative tone in Bantu. In Shigeki Kaji (ed.), *Cross-linguistic studies of tonal phenomena*, 237–257. Tokyo: Institute for the Study of Languages and Cultures.

Hyman, Larry M. & Kemmonye C. Monaka. 2011. Tonal and non-tonal intonation in Shekgalagari. In Sónia Frota, Gorka Elordieta & Pilar Prieto (eds.), *Prosodic categories: Production, perception and comprehension*, 267–289. Dordrecht: Springer.

Jokweni, Mbulelo Wilson. 1995. *Aspects of IsiXhosa phrasal phonology*. Ph.D. dissertation, University of Illinois at Urbana–Champaign.

Kanerva, Jonni M. 1990. *Focus and phrasing in Chichewa phonology*. New York: Garland.

Khoali, Benjamin T. 1991. *A Sesotho tonal grammar*. Ph.D. dissertation, University of Illinois at Urbana–Champaign.

Kisseberth, Charles W. & David Odden. 2003. Tone. In Derek Nurse & Gerard Philippson (eds.), *The Bantu languages*, 59–70. London: Routledge.

Kunene, Daniel P. 1972. A preliminary study of downstepping in Southern Sotho. *African Studies* 31(1), 1–24.

Landis, J. Richard & Gary G. Koch 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.

Leben, William R. 1973. *Suprasegmental phonology*. Ph.D. dissertation, MIT.

Lehiste, Ilse & Gordon E. Peterson. 1961. Some basic considerations in the analysis of intonation. *The Journal of the Acoustical Society of America* 33, 419–425.

Liberman, Mark, Michael J. Schultz, Soonhyun Hong & Vincent Okeke 1993. The phonetic interpretation of tone in Igbo. *Phonetica* 50(3), 147–160.

Lombard, Daan P. 1976. *Aspekte van toon in Noord-Sotho* [Aspects of tone in Northern Sotho]. Ph.D. dissertation, University of South Africa.

Marlo, Michael R. & David Odden. 2019. Tone. In Mark van de Velde, Koen Bostoen, Derek Nurse & Gerard Philippson (eds.), *Bantu languages*, 2nd edn., 150–171. London & NewYork: Routledge.

McCarthy, John J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17, 207–263.

Mmusi, Sheila O. 1992. *Obligatory contour principle effects and violations: The case of Setswana verbal tone*. Ph.D. dissertation, University of Illinois at Urbana–Champaign.

Monareng, William M. 1992. *A domain-based approach to Northern Sotho tonology: A Setswapo dialect*. Ph.D. dissertation, University of Illinois at Urbana–Champaign.

Myers, Scott. 1997. OCP effects in Optimality Theory. *Natural Language & Linguistic Theory* 15(4), 847–892.

Myers, Scott. 1998. Surface underspecification of tone in Chichewa. *Phonology* 15, 367–391.

Myers, Scott. 1999. Tone association and F0 timing in Chichewa. *Studies in African Linguistics* 28(2), 215–239.

Myers, Scott. 2003. F0 timing in Kinyarwanda. *Phonetica* 60, 71–97.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.

Odden, David. 1980. Associative tone in Shona. *Journal of Linguistic Research* 1(2), 37–51.

Odden, David. 1982. Tonal phenomena in Kishambaa. *Studies in African Linguistics* 13, 177–208.

Odden, David. 1986. On the role of the Obligatory Contour Principle in phonological theory. *Language* 62(2), 353–383.

Selkirk, Elisabeth O. 1986. On derived domains in sentence phonology. *Phonology* 3, 371–405.

Selkirk, Elisabeth O. 2011. The syntax–phonology interface. In John Goldsmith, Jason Riggle & Alan Yu (eds.), *The handbook of phonological theory*, 435–484. Oxford: Blackwell.

Snider, Keith L. 1998. Phonetic realisation of downstep in Bimoba. *Phonology* 15, 77–101.

Snider, Keith L. 2007. Automatic and nonautomatic downstep in Chumburung: An instrumental comparison. *Journal of West African Languages* 34, 105–114.

Traill, A., J. Khumalo & P. Fridjhon. 1987. Depressing facts about Zulu. *African Studies* 48, 255–274.

Truckenbrodt, Hubert. 1995. *Phonological phrases: Their relation to syntax, focus, and prominence*. Ph.D. dissertation, MIT.

Turk, Alice, Satsuki Nakai & Mariko Sugahara. 2006. Acoustic segment durations in prosodic research: A practical guide. In Stefan Sudhoff, Denisa Lenertová, Roland Myer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter & Johannes Schließer (eds.), *Methods in empirical prosody research*, 1–27. Berlin: De Gruyter.

Whalen, Douglas H. & Andrea G. Levitt. 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23, 349–366.

Yip, Moira. 1988. The obligatory contour principle and phonological rules: A loss of identity. *Linguistic Inquiry* 19(1), 65–100.

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

Zerbian, Sabine. 2007. Phonological phrasing in Northern Sotho. *The Linguistic Review* 24, 233–262.

Zerbian, Sabine & Etienne Barnard. 2009. Realisation of a single high tone in Northern Sotho. *Southern African Linguistics and Applied Language Studies* 27(4), 357–380.

Zerbian, Sabine & Etienne Barnard. 2010. Realisation of two adjacent high tones: Acoustic evidence from Northern Sotho. *Southern African Linguistics and Applied Language Studies* 28(2), 101–121.

Ziervogel, Daniel, Daan P. Lombard & P. C. Mokgokong. 1969. *A handbook of the Northern Sotho language*. Pretoria: Schaik.