

DOWNSTEP IN TSWANA (SOUTHERN BANTU)

Sabine Zerbian¹ & Frank Kügler²

¹ University of Stuttgart, ² University of Potsdam
sabine.zerbian@ifla.uni-stuttgart.de, kuegler@uni-potsdam.de

ABSTRACT

The article investigates the realization of adjacent high tones across word boundaries in Tswana, a Southern Bantu language. The results show that downstep, a lowering of the second in a series of adjacent high tones, takes place across word boundaries within the same phonological phrase. Downstep does not occur across phonological phrase boundaries. The study confirms an empirical phenomenon and a phonological analysis previously suggested for Southern Sotho for the Tswana variety under consideration by providing quantitative details on the phonetic implementation of this prosodic phenomenon. Quantitative data is rarely presented for tonal phenomena in Bantu languages.

Keywords: downstep, tone, Bantu languages, phonological phrases, OCP

1. INTRODUCTION

Tswana belongs to the Sotho-Tswana group of Southern Bantu languages (S30 in [8]), together with Southern Sotho (Sesotho), Northern Sotho (Sepedi) and Lozi (Silozi). Several descriptions are available in the linguistic literature on aspects of the tonal system of these languages (see [20] for overview). As in other Bantu languages, tone has a lexical and a grammatical function in Sotho-Tswana, and Tswana has two surface level tones, high (H) and low (L).

Bantu languages are best known for their tonal mobility (see [18: 66]) and tone sandhi. Tone sandhi, i.e., tonal changes due to neighbouring tones, can be observed at the word and phrase level. Phonologically, only high tones are assumed to be present in most Bantu languages with low tones being the default tones which are inserted late in the derivation [18]. High tones can be observed to be the active tones, taking part in tone shift, spread, deletion, and/or fusion.

Previous research on Tswana has shown that this language obeys the Obligatory Contour Principle (OCP; [13]) which prohibits identical adjacent autosegments. This restriction holds at the word-level and results in one of three processes: either the blocking of high tone spread if that would create a violation of the OCP, the fusion of two singly-linked tones into one or the delinking of a high tone [15].

For the closely related and mutually intelligible language Southern Sotho, Khoali [11] shows the OCP to be relevant not only at the word-level but also at the phrase-level. In a sequence of two adjacent high tones across word boundaries, a word initial high tone is downstepped in Southern Sotho if both words occur within the same Phonological Phrase. If the two high tones are adjacent across an intervening phrase boundary, no downstep occurs (for more details see section 2).

Creissels [5] reports downstep at the phrase-level also for Tswana. His approach is not phrase-based though, and cannot be considered in the scope of this article.

Khoali [11] and Creissels [5] explicitly state that dialectal variation is to be expected across the different varieties of Sotho-Tswana, not surprisingly also in the tone system. Creissels [5] even assumes that responses to violations of the OCP are different across varieties of Tswana.

The article presents a study that set out

- to test in a controlled way whether Tswana realizes adjacent high tones across word boundaries within a Phonological Phrase by means of downstep, thereby following the pattern reported for Southern Sotho
- to provide quantitative data on the realisation of downstep in a Southern African Bantu language
- to formulate a generalization for the occurrence of downstep in Tswana.

The article is structured as follows: Section 2 presents the basic insights of Khoali's work on downstep in Southern Sotho, which informed the selection of stimuli for the study on Tswana. Section 3 reports methodological details. Section 4 presents the results. The article closes with a discussion of the results and a sketch of the phonological analysis in section 5.

2. DOWNSTEP IN SOUTHERN SOTHO

Khoali [11] reports that in a sequence of two adjacent high tones across word boundaries, a word initial high tone is downstepped in Southern Sotho if both words occur within the same Phonological Phrase (1a) (marked by an exclamation mark). If the two high tones are adjacent across an intervening phrase boundary, no downstep occurs (1b).

- (1) Downstep in S. Sotho [11: 65, 114]
- a. (Pódí !é-rátá moríti)ϕ
 NP9.goat SC9-like NP3.shade
 ‘A goat likes shade.’
- b. (Di-pódi)ϕ (tsé-!kgóló hahóló !dí-fihlile)ϕ
 NP10-goat POSS10-big very SC10-arrived
 ‘Very big goats have arrived.’ (*Dipódí !tsé)

This approach is consistent with Kunene’s [12] descriptive pilot study on downstep in Southern Sotho.

Khoali [11] follows an indirect reference approach to the syntax-phonology interface which sees phrasing being indirectly related to syntax: A subject and a following verb constitute one Phonological Phrase. A head noun is separated from its modifier by a Phonological Phrase boundary, as is the case in the possessive construction in (1b).

Further language-internal and external evidence confirms the phrasing in (1). For instance, disyllabic nouns with an underlying high tone on the stem initial syllable (*pó-* in (1)) are realized with an HL-pattern when they occur final in a Phonological Phrase, as in (1b). Thus, the occurring high tone spread to the final syllable of the disyllabic subject in (1a) marks this constituent as not phrase-final [11: 70]. Also across Bantu languages, it has been reported that the subject constitutes one Phonological Phrase with the verb (e.g., [2, 9, 10, 19]).

3. METHODOLOGY

3.1 Stimuli

In order to test whether syntactic context and hence phrasal structure makes a difference in the realization of two adjacent high tones across word boundaries also in Tswana, and if so, how the difference manifests itself phonetically, a production study was designed. Target sentences were constructed in such a way that two adjacent high tones occur across word boundaries in the syntactic contexts exemplified in (1). A corresponding example from Tswana is given in (2).

- (2) Target sentence from Tswana
- a. Baná bá-tshábá makgóa.
 NP2.child SC2-fear NP6.white_pers.
 ‘The children fear the White people.’
- b. Ntšá yágágó é-jélé dijó tsámé.
 NP9.dog POSS.you SC9-ate NP8.food POSS.I
 ‘Your dog has eaten my food.’

Context A, (2a), shows a high tone in the final syllable of a subject noun being followed by a verb

with an initial high-toned subject concord and a following high-toned verb stem. Context C, (2b), showed a high tone in the final syllable of a noun followed by a high-toned possessive concord of a following nominal modification. The possessive concord is followed by a high-toned noun to meet the sequence of three consecutive high tones, parallel to context A. Four different target sentences were constructed for each of these two contexts. The total set of test sentences comprised three repetitions of the 4 sentences per context.

3.2 Recording procedure

The 24 sentences (2 contexts x 4 sentences x 3 repetitions) were presented in pseudo-randomized order using presentation software, interspersed with filler items from another study. 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 the sentences first to familiarize themselves with the sentences and the recording setting. 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. Although care was taken to minimize background noise, it could not be fully avoided. Data were transferred to a computer hard drive for further analysis.

3.3 Participants

Twelve speakers (5 male, 7 female, aged 15 - ~50) took part in the study. Speakers were selected by the local pastor because their speech was considered representative for the local area. According to Cole [4: xvi], the variety of Tswana spoken in this region belongs to the Southern division of Tswana.

3.4. Analysis

The main acoustic correlate of tone in Bantu languages is fundamental frequency (F0) [15], which can only be realized on voiced segments. In order to guarantee comparability across all stimulus items, only the vowels of the high-toned syllables were delineated and analysed for their F0. 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 (H1 and H2 respectively).

The mean of the central 50% of the tone-bearing vowel was used as acoustic reference for the tone of the vowel in order to cut off microprosodic influences by preceding and/or following segments (which could not be controlled for).

The presence or absence of downstep is then established through a comparison between the F0 of H1 and H2 (see [16, 17] for downstep in Bimoba and Chumburung, [6, 7] for Akan). Downstep occurs when H2 is systematically lower than H1.

4. RESULTS

4.1. Adjacent high tones within a phrase

In context A, downstep has been reported to occur between the final high tone of the first word (H1) and the first high tone of the second word (H2) even in Tswana [5]. H1 and H2 are thus expected to differ significantly, which means that downstep is realized.

A series of paired samples t-tests were run examining for each speaker whether H1 and H2 were significantly different in Context A. We found diverging results (see Table 1): Seven out of 12 speakers realized a significantly downstepped H2. One speaker (06) very closely approached the significance level of $p < 0.05$ and was therefore included in the further analysis. For two speakers (04, 09) the difference between H1 and H2 in context A is not significant. Two other speakers (05, 07) show a significant effect, though in the other direction (indicated by negative t -values).

Table 1: Estimation of downstep realization between H1 and H2 for each speaker in context A. Significance levels: * 0.05, ** 0.01, *** 0.001.

	sex	age	H1 (Hz)	H2 (Hz)	diff.	Statistics
01	F	~50	228	194	***	$t = 7.8, df=11, p < 0.001$
02	M	~20	99	94	**	$t = 4.1, df=9, p < 0.01$
03	F	~22	208	199	***	$t = 6.3, df=11, p < 0.001$
04	M	~30	145	143	n.s.	$t = 1.6, df=11, p = 0.1371$
05	M	~17	136	141	**	$t = -3.8, df=11, p < 0.01$
06	F	~15	242	233	n.s.	$t = 2.2, df=10, p = 0.05337$
07	M	~45	126	130	**	$t = -3.9, df = 9, p < 0.01$
08	M	~35	126	110	***	$t = 8.5, df = 11, p < 0.001$
09	F	~20	225	221	n.s.	$t = 1.2, df=10, p = 0.274$
10	F	~50	258	242	***	$t = 6.1, df = 9, p < 0.001$
11	F	~22	244	230	*	$t = 2.4, df = 11, p < 0.05$

Given our definition of downstep above, 8 speakers realize downstep in this context, whereas 4 speakers do not. We decided not to pool the results across these two groups as this might average out the two different patterns we observe.¹

We continued the analysis with the data of those 8 speakers who did realize downstep in context A. Items were excluded which contained hesitations, mispronunciations, or pauses affecting

the syllables under consideration, or a considerable amount of creaky voice which does not allow F0 measures (resulting in 90 target items for context A and 84 in context C). The aim was to investigate whether downstep is produced whenever two adjacent high tones occur across word boundaries, or whether its occurrence depends on the syntactic and thus phrasal context, as has been reported for Southern Sotho.

4.2. Comparison of adjacent high tones within a phrase and across a phrase boundary

4.2.1 Representative illustrations

Figure 1 and 2 exemplify the tonal realization in context A and C. In each figure, the oscillogram, the pitch track, orthographic words and target vowels are shown.

Figure 1: Context A, sentence 2, speaker 01

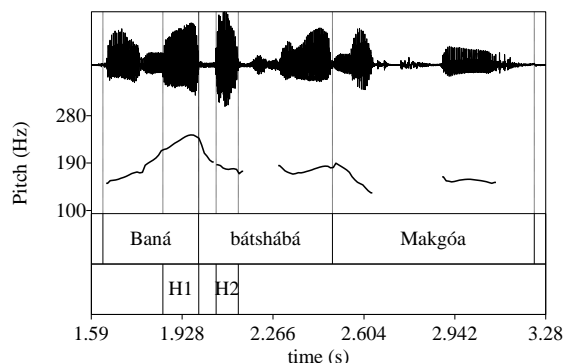


Figure 2: Context C, sentence 4, speaker 07

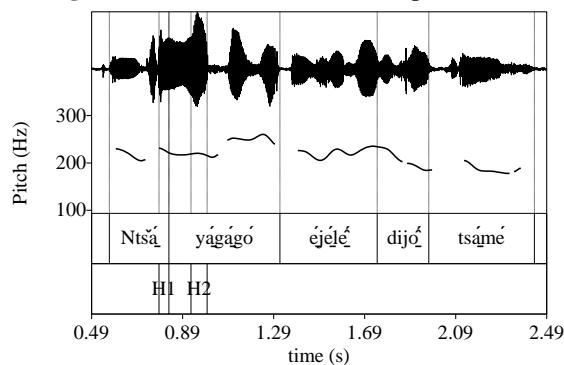


Figure 1 represents context A where we see in the pitch track that H1 is realized at a higher pitch than H2. Figure 2 represents context C. We see in the pitch track that H1 and H2 are realized at roughly equal pitch heights.

4.2.2. Descriptive statistics

Figure 3 shows the mean values of H1 and H2 respectively in the two contexts, averaged across all eight speakers (comprising male and female). The

black dots give the average mean F0, the lines the 95% confidence interval.

Figure 3: Averaged means for H1 and H2 across all 8 speakers split by contexts

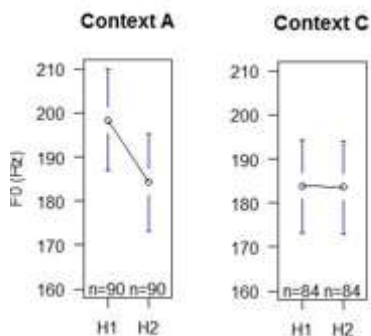


Figure 3 shows that, on average, there is a considerable drop in F0 from H1 to H2 in context A whereas F0 stays level in context C.

4.2.3. Inferential statistics

The visually evident difference between H1 and H2 in contexts A and C was also tested for its statistical significance. The values of H1 and H2 in each context are compared by means of a paired samples *t*-test to account for the differences in pitch height between women and men. The results show a significant difference between H1 and H2 in context A ($t = 9.6$, $df = 89$, $p < 0.001$). No significant difference emerges between H1 and H2 in context C ($t = 0.3$, $df = 83$, $p = 0.803$). The results are thus consistent with the patterns shown in figure 3, and show that downstep, i.e., a phonetic lowering of an underlying high tone, takes place in context A but not in context C.

5. ANALYSIS AND DISCUSSION

The results confirmed for Tswana, a variety of the Sotho-Tswana group of Southern Bantu languages, the claim that has been made in the literature for Southern Sotho, namely that the occurrence of downstep above word-level depends on syntactic structure. More precisely, downstep takes place between adjacent high tones belonging to a subject and a verb (2a), but not between high tones belonging to a noun and a following modifier (2b).

Given the close relatedness of Southern Sotho and Tswana and their parallel behaviour concerning the occurrence of downstep, it is natural to adopt the phrasal analysis of Southern Sotho [11] for the Tswana variety under consideration. Thus, downstep occurs on two adjacent high tones across words within a Phonological Phrase, whereas it is absent if the two high tones are separated by word- and phrase boundaries.

Khoali [11: 17] captures this generalization by means of a phrasing algorithm which defines that the phonological phrase “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”.

Like in Southern Sotho this phrasing is supported by a parallel tonal rule: The last syllable of a Phonological Phrase is exempt as a target for High Tone Spread in Tswana (see [3] for data supporting this).

In his work on the Zambian Bantu language Namwanga, Bickmore [1] analyses downstep as the phonetic interpretation of a phonological OCP violation. Namwanga “tolerates structural violations of the OCP in the output of phonology. When the phonetic component encounters two consecutive TBUs linked to distinct H’s, a downstep is realized between them” [1: 302]. In Tswana, however, the surface contrast shows that downstep is sensitive to prosodic phrasing too in this language. Thus, downstep cannot merely be a matter of phonetic implementation in this variety of Tswana. The emerging surface contrast between downstepped and non-downstepped sequences of high tones are thus mirrored in phonological representation.

6. REFERENCES

- [1] Bickmore, L.S. 2000. Downstep and fusion in Namwanga. *Phonology* 17, 297-331.
- [2] Cheng, L., Downing, L.J. 2009. Where’s the topic in Zulu? In De Hoop, H., Van Bergen, G. (eds.) Special issue: *Topics cross-linguistically*. *The Linguistic Review* 26, 207-238.
- [3] Cole, D.T., Mokaila, D.M. 1962. *A Course in Tswana*. Washington: Georgetown University.
- [4] Cole, D.T. 1955. *An Introduction to Tswana Grammar*. Cape Town: Longman.
- [5] Creissels, D. 1998. Expansion and Retraction of High Tone Domains in Setswana. In: Hyman, L., Kisseberth, C. (eds), *Theoretical aspects of Bantu tone*. Stanford, Center for the Study of Language and Information: 133-194.
- [6] Dophyne, F.A. 1994. A phonetic and phonological study of downdrift and downstep in Akan. Paper presented at the 25th Annual Conference on African Linguistics, Rutgers University.
- [7] Genzel, S., Kügler, F. 2011. Phonetic realization of automatic (downdrift) and non-automatic downstep in Akan. *Proc. 17th ICPhS*, Hong Kong, 735-738.
- [8] Guthrie, M. 1967-1971. *Comparative Bantu: An Introduction to the Comparative Linguistics and*

Prehistory of the Bantu Languages. (Volumes 1-4).
Farnborough: Gregg International.

- [9] Jokweni, M.W. 1995. Aspects of IsiXhosa Phrasal Phonology. PhD dissertation. University of Illinois at Urbana-Champaign.
- [10] Kanerva, J.M. 1990. *Focus and Phrasing in Chichewa Phonology*. New York: Garland.
- [11] Khoali, B.T. 1991. A Sesotho tonal grammar. PhD thesis. University of Illinois at Urbana-Champaign.
- [12] Kunene, D.P. 1972. A preliminary study of downstepping in Southern Sotho. *African Studies* 31:1, 1-24.
- [13] Leben, W.R. 1973. Suprasegmental phonology. PhD thesis, Massachusetts Institute of Technology, Massachusetts.
- [14] Mmusi, S.O. 1992. Obligatory contour principle effects and violations: the case of Setswana verbal tone. PhD thesis, University of Illinois at Urbana-Champaign.
- [15] Myers, S. 1998. Surface underspecification of tone in Chichewa. *Phonology* 15, 367-391.
- [16] Snider, K.L. 1998. Phonetic realisation of downstep in Bimoba. *Phonology* 15, 77-101.
- [17] Snider, K.L. 2007. Automatic and nonautomatic downstep in Chumburung: an instrumental comparison. *Journal of West African Languages* 34, 105-114.
- [18] Yip, M. 2002. *Tone*. Cambridge: Cambridge University Press.
- [19] Zerbian, S. 2007. Phonological Phrasing in Northern Sotho. *The Linguistic Review* 24, 233- 262.
- [20] Zerbian, S., Barnard, E. 2009. Realisation of a single high tone in Northern Sotho. *Southern African Linguistics and Applied Language Studies* 27(4), 357-380.

ⁱ An explorative investigation into the realization of adjacent high tones in context C by those four speakers who did not realize a downstep in context A revealed no downstep in context C. One can only speculate why they do not realize a downstep in context A. Little is known about inter-speaker variation in tone.