

Realisations of a single high tone in Northern Sotho

Sabine Zerbian^{1*} and Etienne Barnard²

¹*Department of Linguistics, University of the Witwatersrand, Private Bag 3, WITS 2050, South Africa*

²*Meraka Institute, Human Language Technologies Research Group, Council for Scientific and Industrial Research, PO Box 395, Pretoria 0001, South Africa*

**Corresponding author, e-mail: sabine.zerbian@gmail.com*

Abstract: This article reports on a production study that investigates the realisation of a single high tone in the verbal constituent in Northern Sotho, a Bantu language spoken in South Africa. The parameters of variation investigated are based on existing descriptive and theoretical literature and relate to numbers of syllables in the verb stem, morphosyntactic constituency and verb-internal morphological boundaries. The results are interpreted as both phonological and phonetic influences on high tone realisation in this language: phonologically, a high-toned object concord causes a peak shift one syllable to the right. Phonetically, the study shows that the F0 peak associated with a high tone is not necessarily reached within the syllable carrying the high tone but only later (*peak delay*) depending on the segmental make-up of the tone-bearing syllable and its position within the utterance. The segmental make-up of the tone-bearing syllable leads to systematic surface variation in tone realisation. By collecting controlled acoustic data on tone realisation, this study provides a ground for cross-dialectal comparison of Southern Bantu tone.

Introduction

The current study reinvestigates the phonetic realisation of a single high tone in Northern Sotho, a Southern Bantu language belonging to the Sotho-Tswana group. It analyses acoustic data from four speakers of the same age group¹ using stimuli which are controlled for interfering segmental influences. Parameters of variation within the stimuli set are: tense-mood-aspect, length of the verb stem (ranging from disyllabic to quattrosyllabic verb stems), and morphological origin of the single high tone (ranging from subject concord, tense marker, mood marker, object concord to verb stem). These parameters have been reported in the literature to influence the realisation of high tones in Bantu languages in general.

The current study does thus not aim at providing an exhaustive investigation into the overall tone system of Northern Sotho. Coherent descriptions are available by Lombard (1976) and Ziervogel *et al.* (1969), as summarised in Table 1 below. The aim of the present study is rather the investigation into the detailed realisation of one single high tone in the verbal domain, factoring out potential interferences with respect to stimuli, speakers or transcribers. This is done through controlling for segmental and grammatical contexts, using acoustic data, and collecting several repetitions as well as several speakers, which nevertheless form a homogeneous group with respect to geographical origin and age. Trends observable in all speakers are the main focus of our interest, as those are the ones that are most likely characteristic of the variety as a whole. Despite our empirical approach, ambiguities in the interpretation of these data remain, and we try to give reasonable interpretations in those cases.

The data collected in the current study are transcribed using the familiar, categorical notation of high and low tones. When comparing the transcriptions resulting from the interpretation of the acoustic data to the existing descriptions of corresponding structures in other Sotho-Tswana varieties (see the next section), it emerges that there is no perfect match to any reported variety. Instead of postulating a new variety with unique tonal features, the study is meant to advocate the methodology used and to encourage dialectal studies into tone that gather data that can be compared across dialects, speakers and transcribers.

However, besides describing the tonal patterns found in the four speakers who participated in our study, the current article also differentiates between the phonological and phonetic factors that interact in the surface realisation of tone in these contexts in Northern Sotho.

Whereas the exploration of the phonetic-phonology interface by means of acoustic studies is not a new methodology as such, it is hardly ever applied in Bantu tonology (with the notable exception of Myers, 1998b, 1999, 2003 on Chichewa and Kinyarwanda).

In the current study morphosyntactic constituency and length of the verb stem are investigated as the independent variables in high tone realisation based on findings from the theoretical literature. Previous acoustic research has concentrated on high tone alignment in relation to syllable length (Myers, 1999, 2003). In the current study, syllable length is kept constant.

The remainder of the article is structured as follows: the next section summarises tone in Sotho-Tswana by reviewing the existing tonal descriptions with respect to the constructions investigated in the current study. The section following then describes the production study that was carried out in order to investigate the phonetic realisation of high tones in the speech of four speakers of Northern Sotho. The results are then presented; and the final section discusses the results both with respect to the phonetics-phonology interface of tone as well as with specific reference to Northern Sotho.

Tone in Sotho-Tswana

This section lays out the phonological and phonetic characteristics of the Sotho-Tswana tone system that pertain to the alignment of a single, phrase-medial high tone that is surrounded by low tones. The realisation of high tones which are adjacent to other high tones shows different characteristics and is dealt with in a separate study for Northern Sotho (Zerbian & Barnard, unpublished data).

Basic characteristics

The languages of the Sotho-Tswana group share the basic tonal characteristics with the majority of Bantu languages. They use tone to distinguish grammatical and lexical meaning. Their tone inventory consists of two tones, H and L. The high tone (H) is the active tone in the languages of the Sotho-Tswana group as it participates in tone spread and is subject to positional restrictions, such as the avoidance of two adjacent high tones immediately next to each other (so-called OCP context) and the banning of phrase-final high tones (see Kisseberth & Odden, 2003 for a general discussion of these aspects). Low tones (L) can appear next to each other without inducing any tonal changes and they also appear on phrase-final syllables.

The current study focuses on the realisation of high tones because high tones are the ones that are assumed to be specified underlyingly whereas low tones are considered to be either inserted late in the phonological derivation or only implemented phonetically. The view of underspecification of low tones has found phonetic support in Myers' (1998b) acoustic study of low tones in Chichewa, a Bantu language spoken in Malawi. As high tones are specified in phonological representation they have a phonetic pitch target. Low tones, being underspecified phonologically, lack a phonetic pitch target and can be derived via (sagging) transitions between high tones (Myers, 1998b; but see Xu & Xu, 2005 on English intonation; Chen & Xu, 2006 against sagging transition as a viable articulatory mechanism).

Similarly to other Bantu languages, the languages of the Sotho-Tswana group show an asymmetry in the tonal characteristics of their noun and verb system with nouns being more tonal than verbs. However, Bantu languages are agglutinative languages and can thus have a complex, extended verb word in which tones interact with each other morphologically. The realisation of a single high tone within the morphologically complex verb word is the focus of the current study.

Determining factors in tone alignment

There are various parameters that are known from the literature to influence (high) tone realisation in Bantu languages. These are grammatical aspects such as polarity (negative vs positive) and tense-aspect-mood; pragmatic aspects such as focus and sentence type; prosodic aspects such

as vowel length and size of the tonal domain; and other aspects such as position within a phrase and morphosyntactic constituency. In the current study, only tense-aspect-mood, size of the tonal domain, and morphosyntactic constituency have been varied in a controlled way.

All other parameters have been kept constant, as they did not constitute the focus of the current study. Polarity has been kept constant by using only positive target sentences; sentence type by using only positive statements; focus by only using un-focused verb words (Northern Sotho uses non-prosodic means of focusing, see Zerbian, 2006); and position within a phrase by having the target word appear in phrase-medial position via an object or adverb following it. As for vowel length, Northern Sotho does not have phonological vowel length.

Only the three parameters relevant for the current study are discussed in the following.

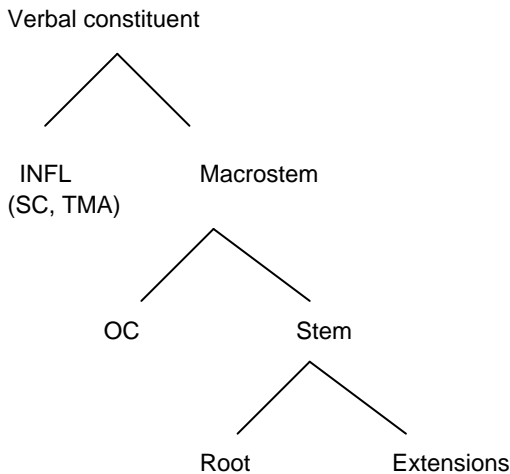
Morphosyntactic constituency

The verb in Bantu languages shows a rich agglutinative structure, which can be described by the template in (1), where SC and OC refer to subject and object concord markers respectively, and Ext refers to verbal extensions such as causative, applicative or reciprocal. Optional elements are given in brackets.

(1) SC – (Tense, Mood, Aspect) – (OC) – verb root – (Ext)

Much work (e.g. Barrett-Keach, 1986; Myers, 1987, 1998a) has argued that the Bantu verbal constituent is ‘split’ into two distinct morpho-phonological constituents: the Inflectional stem, comprising subject concord marker (SC) and tense, mood, and aspect prefixes (TMA); and the Macrostem, comprising the object concord marker (OC) and the verb stem with extensions. The proposed structure is shown in (2).

(2) Bantu verbal structure (adapted from Myers, 1987; Downing, 1999)



The Stem consists minimally of the verb root ending in a final suffix. Derivational suffixes can occur between root and final suffix. The suffix structure is captured in somewhat simplified fashion by ‘extensions’ in (2) as these do not form the focus of attention in the present article. If present, the object concord forms a morpho-phonological constituent with the verb root, termed Macrostem following Hyman & Ngunga (1994).

Without going into the morphosyntactic reasoning for adopting such a morphologically split verbal structure (see e.g. Barrett-Keach, 1986; Myers, 1987; Hyman, 1993; Mchombo, 1993; Mutaka, 1994 for evidence from various Bantu languages), the morpho-phonological reasoning for this split-constituent hypothesis is important for the realisation of tone. The stem is commonly subject to prosodic minimality restrictions. It constitutes the domain for vowel harmony, nasal consonant harmony, and restrictions on consonant and vowel distribution (Downing, 1999, 2001). Furthermore, there is tonal evidence that the morphological constituents INFL, Macrostem, and Stem also

constitute distinct tonal domains as tonal processes such as Meeussen's rule and high tone spread make reference to these domains (e.g. Myers; 1987; Hyman & Mathangwane, 1998; Downing, 2001). The examples in the literature are transcribed data as acoustic data are not yet available.

Tonal evidence for the verb subconstituents is illustrated in (3) with data from Durban Zulu where a high tone from a subject concord targets the antepenultimate syllable for its rightmost edge of realisation, (3a), whereas a stem-initial high tone targets the penultimate syllable, (3b). Further evidence for the relevance of the morphological structure for tone in Durban Zulu is that a pre-stem high tone spreads, (3a), whereas a stem-initial high tone shifts, (3b). Lexical high tones are underlined, realized high tones are marked by acute accent.

(3a) úkú-kákísana 'to cause each other to surround'

(3b) si-ya-thengiselána 'we sell to each other'

Durban Zulu (Downing, 2001: 2)

Morphological constituency as in (2) can thus exert an influence on tone, in two possible respects: first, with respect to an immediately following morphological boundary, and second with respect to which morphological constituent a high tone originates from. Both parameters have been reported as significant for tone realisation in the existing literature on Southern Bantu languages and are thus controlled for in the current study (see the final section for discussion).

Tense, mood and aspect

In Sotho-Tswana, as in Bantu languages more generally, differentiations in tense, mood and aspect are expressed via morphemes or a combination of a morpheme and a specific tone pattern (so-called *tone melodies*).²

If TMA is expressed solely via a morpheme, then that morpheme will come with its tone, either high or low. The tonal surface realisation of the entire verb word will be according to more general tonal rules in that language. Examples from Tswana are the morphemes indicating future tense and a persisting action. The future marker *tlo* is underlyingly low-toned (or underspecified). Its surface tone is reported to follow more general tonal rules in Tswana, as illustrated in the examples in (4) (Cole & Mokaila, 1962). If all syllables in the verb word are low-toned (or unspecified for tone), the surface realisation will be all low, unsurprisingly, (4a). If the verb stem which follows the future marker *tlo* is high-toned, the verb stem initial syllable will be high, (4b). If only the subject concord is high-toned, (4c), the future marker *tlo* will be realised with a high tone, due to high tone spread. Finally, if both the subject concord and the verb stem are high-toned, the future marker *tlo* will retain its low tone due to the blocking of high tone spread if the following syllable is itself high-toned.

(4a) ke-tlo-hlaba... 'I will slaughter...'

(4b) ke-tlo-r^éká... 'I will buy...'

(4c) ó-tló-hlaba... 'S/he will slaughter...'

(4d) ó- tlo-r^éká... 'S/he will buy...'

Thus, the tonal surface realisation in the future tense can be derived by the underlying tone of the future morpheme in conjunction with general tone rules of the language. The same holds for the persistive marker *sá*, except that this marker is itself high-toned, thus a different set of general tone rules will apply. If the high-toned persistive marker occurs in an all-low environment, high tone spread is reported to occur, (5a). If it is preceded and/or followed by a high tone, all high tones are reported to be realised, (5b–d) (for a discussion of the realisation of two adjacent high tones, see Zerbian & Barnard, unpublished data).

(5a) ke-sá-hlába... 'I am still slaughtering...'

(5b) ke-sá-r^éká... 'I am still buying...'

(5c) ó-sá-hlaba... 'S/he is still slaughtering...'

(5d) ó-sá-r^éká... 'S/he is still buying...'

Again, the tonal surface realisation can be derived from the underlying tone of the morpheme in conjunction with general tonal rules of the language. Thus, the TMA category is not reported to influence the tone pattern beyond the effect of the underlying tones.

This is in contrast to other tenses which are not only marked by a morpheme but also by a special tone pattern. One example is the perfect tense which is morphologically marked by the extension

-ilê (or one of its allomorphs). In addition, a verb form in the perfect tense surfaces with a distinct tone pattern, characterised by a sequence of high tones even for underlyingly low-toned (unspecified) verbs. This is exemplified in (6).

- (6a) ke-apéílé... 'I cooked...'
 (6b) ke-rékílé... 'I bought...'
 (6c) ó-apéílé... 'S/he cooked...'
 (6d) ó-rékílé... 'S/he bought...'

One can assume that the perfect tense is characterised by morphemes and a specific tone melody (Cole & Mokaila, 1962; see Khoali, 1991 for an alternative analysis in which the grammatical tone links to the second stem syllable and spreads iteratively).

The current study investigates the realisation of a single high tone only in morphological contexts such as (4) and (5), where the surface tone can be derived by the underlying tone and general tonal rules of the language.

Size of the tonal domain

Various tonal descriptions of Southern Bantu languages note an alternation in the target of high tone spread depending on the length of the verb stem. Monareng (1992) in his study on Setswapo, a variety closely related to Northern Sotho, reports that a high tone spreads to the penultimate syllable in disyllabic verb stems whereas it spreads to the antepenultimate syllable in longer stems.

Similar observations are reported for the Nguni languages. Here, the target for high tone realisation is the antepenultimate syllable, (7b), except for disyllabic and trisyllabic verb stems for which the penultimate syllable is targeted for high tone realisation, (7a).

- (7a) si-ya-bóna 'we see'
 si-ya-fumána 'we get'
 (7b) si-ya-sebénzisa 'we are working'

Xhosa (Cassimjee & Kisseberth, 1998)

The number of syllables in the verb stem is thus potentially relevant for high tone realisation as the reports from the literature on tone in Southern Bantu languages show (note, however, that in the examples cited above the verb always appears in phrase-final position whereas in our study it appears in phrase-medial position, indicated by '...').

Dialectal differences in high tone realisation

The Sotho-Tswana group (S30) comprises the four languages Tswana, Southern Sotho, Northern Sotho and Lozi. They are spoken in Botswana, Lesotho, South Africa and Zambia. All these languages are mutually intelligible and they will thus be considered varieties of each other for the purpose of this article.

Northern Sotho is the written language of the Limpopo Province and Mpumalanga. It was originally based on the Pedi variety and later had elements from Kôpa incorporated. Subsequently, the dialects from around present-day Mokopane, namely Moletši, Matlala and Tlôkwa affected the written form so that today's written Northern Sotho can no longer be considered representative of Pedi (Ziervogel *et al.*, 1969: 1).

Several descriptions of the tone systems of the Sotho-Tswana languages are available. In Table 1, only descriptions are listed which investigate a major part of the tone system, either in total, or a considerable part of the verbal morphology.

The descriptions agree in one basic property which can thus be assumed to be characteristic of tone in the Sotho-Tswana group, namely the fact that tone spreads to the right-adjacent syllable if the tonal context allows it. Examining the existing descriptions of the Sotho-Tswana group with regard to the contexts of interest in the current study, we find divergent reports as to the realisation of a single high tone. An overview is given in Table 2. In the presentation of the data, the widespread practice in Bantu linguistics is maintained by marking high tones by acute, underlying high tones by underlining and leaving low tones unmarked.

A comparison of the tone patterns across the different Sotho-Tswana varieties reveals that all Sotho-Tswana varieties are reported to show what is referred to as high tone spread in the

Table 1: Tonal descriptions of the Sotho-Tswana group

Language	Variety	Author
Northern Sotho	Pedi dialect	Lombard (1976)
	Northern Sotho	Ziervogel <i>et al.</i> (1969)
	Tswapo dialect	Monareng (1992)
Tswana	Ngwaketse dialect	Chebanne <i>et al.</i> (1997)
	Rolong dialect, South Africa	Cole & Mokaila (1962)
		Cole (in press)
Southern Sotho	Rolong, Tlhaping	Mmusi (1992)
	East/ South-East Lesotho	Khoali (1991)
	Lesotho, not specified	Lets'eng (1995)

Table 2: Tone patterns of verb forms in Sotho-Tswana varieties

Context	Structure	Northern Sotho Pedi		Southern Sotho	Southern Sotho	Tswana
		Ziervogel <i>et al.</i>	Lombard	Khoali	Lets'eng	Chebanne <i>et al.</i>
1	SC + <u>verb stem</u>	re-rútá re-rútisa (p. 134)	ke-kólóbétša kgóšhi (p. 83)	ke-rékisa nama (p. 239) o-kgúrúmetša motho (p. 190)	ke-bótsá tsila (p. 22) ke-kótúla mabelé (p.22)	ke-thúsá ngáka lo-tshámíkisa metsí (p. 92)
2	<u>SC</u> + verb stem	ó-réta	bá-hlágolela malóme lé-tlíšá dipúdi (p. 84)	ó-bátla nama (p. 205) ó-kgárametša motho (p. 190)	bá-bítsana ká majwe (p. 26)	ó-bátlá gó-tsamaya ó-ápára sesádi (p. 94f)
3	<u>SC</u> + T + verb stem	ó-tló-reta	á-tló-epolla (p. 157)	ó-tlá-tsamaaya (p. 224)	bá-tlá-bitsana ká majwe (p. 72)	bá-tláá-batla ó- tláá-tswela (p. 117f)
4	SC + <u>I</u> + verb stem	re-sá-reta	le-sá-epolla re-sá-apara	not available	ke-sá-bula lematí ke-sá-lokisa kolóí (p. 81)	not available
5	SC + <u>OC</u> + verb stem	re-é-rét re-é-rétela	go-dí-bápala (p. 58)	not available	o-mq-étela mokibelo (p. 25)	ke-ló-bálá naé ke-bá-báléla lokwálo (p. 93f)

phonological literature, i.e. that a lexically underlyingly high tone is not only realised on the syllable it is associated with underlying but also on the immediately right-adjacent syllable. This is only true if no other high-toned syllable follows immediately.

Variety-specific variation can be observed with respect to the domain of high tone spread, though. In Northern Sotho (Ziervogel), Pedi (Lombard) and Southern Sotho (Khoali, and Lets'eng) the domain for high tone spread seems to be only the immediately right-adjacent syllable. In Tswana (Chebanne), tone spread takes place over the next two adjacent syllables. There are exceptions reported to these variety-specific generalisations, some of which are tied to a specific construction, whereas others can be less clearly delineated. Among the less clearly delineated exceptions is Pedi where we also find spread to the following two syllables in the examples for contexts 1 and 2. In Tswana, we find high tone spread to the following three syllables, as in the first example for context 2. These exceptions cannot be delineated on any linguistic ground so far.

Another exception also seems to be tied to a specific construction. From the data available, it seems that there is an interesting, systematic exception to the general pattern of high tone spread which occurs in the persistive verb form in context 4. In this context, the high tone does not seem to spread on to the right-adjacent syllable in any of the varieties. Instead, the high tone is only realised on the tone-bearing syllable.

Another interesting exception occurs in the Southern Sotho variety described by Lets'eng in context 5. Here, the high tone of the object concord is not realised on the tone-bearing morpheme itself but only on the following syllable. This is also what the current study finds for Northern Sotho. In addition, however, the high tone does not spread to further syllables to the right in Southern Sotho.

From the literature review of existing tonal descriptions it remains unclear if morphosyntactic constituency is relevant for tone in Sotho-Tswana, mainly because of diverging reports and incomparable data. Although there might well be dialectal differences between the varieties in exactly this respect, it is difficult to gain a clear picture based on the existing sources. The reason for the differences is, however, difficult to pinpoint. The data given in the literature are mainly based on auditory judgements, which are either based on the author's variety (as in the case of Monareng, Mmusi, Khoali, and Lets'eng) or on the speech of one language consultant (Chebanne *et al.*; Lombard). The empirical basis for the description in Ziervogel *et al.* (1969) is not specified. Is the variation reported indeed dialectal variation due to different geographical origin of the varieties under consideration, are we dealing with speaker idiosyncrasies, or with differences due to the fact that different transcribers transcribed the data?

We thus carried out an acoustic study with four speakers of Northern Sotho in order to collect data that would help determine if length of the verb stem and morphosyntactic constituency are relevant for the realisation of a single high tone in this variety of Sotho-Tswana.

Production study

The production study was carried out in order to gather acoustic data on high tone realisation in Northern Sotho, varying both the morphosyntactic constituent from which the high tone originates as well as the number of syllables of the verb stem.

Stimuli

Table 3 illustrates the varying size of the tonal domain in the context of high tones originating on the verb stem initial syllable. One example is given for each syllable length.

As is marked in Table 3 by underlining, the initial syllable of the verb stems carries a high tone underlyingly.³ Verbs differ with respect to the number of syllables that follow within the verb stem, ranging from one to three. These high-toned verb stems are produced in an all-low-toned environment, i.e. preceded by the low-toned subject concord *re* ('we') and followed by a low-toned noun, either an object or a temporal adverbial depending on the lexical meaning of the verb. A complete list of stimuli with translation can be found in the Appendix. Each token is repeated three times, resulting in 27 utterances per speaker per context.

Table 4 lists the morphosyntactic contexts chosen to determine if high tones are realised differently across morphosyntactic domains or when originating from different morphosyntactic constituents. In the table, the abbreviation SC refers to the subject concord. Depending on tone either *re* 'we' or *ǒ* 's/he' was used. OC refers to the object concord. It turned out in a pilot study that the object concord of class 1 *mo*, which has been reported as the only low-toned object concord. (Lombard, 1976; Ziervogel & Mokgokong, 1979), carries a high tone in the speech of our participants, so that there are no low-toned object markers in the experiment stimuli.⁴

The abbreviation T refers to a tense or aspect marker. In the study, *tlo* has been chosen as the low-toned tense marker. It refers to the future. *Sá* has been chosen as the high-toned aspect marker. It refers to an ongoing action (persistent).

With respect to the segmental make-up of the stimuli, they were generally constructed in such a way to contain sonorants only in order to exclude segmental influence on pitch. However, a few exceptions occurred: one trisyllabic and one quatsyllabic high-toned verb stem were included, starting on a voiceless glottal fricative (see also the final section of this article). Furthermore, there are no tense or aspect markers in Northern Sotho which would satisfy the condition of sonority. The tense and aspect markers that we chose contain a laterally released plosive (<tl>) and a voiceless fricative (<s>). Thus, pitch disturbances due to the segmental influence will be expected to occur in the data.

Table 3: High-toned verb stems

Context	Morpho-phonological context	Morpho-phonological domains	Translation
1	SC + <u>verb stem</u>	[re] _{INFL} -[_{Stem} mêmá]... [re] _{INFL} -[_{Stem} rêmêla]... [re] _{INFL} -[_{Stem} romêlana]...	'we invite...' 'we chop for...' 'we send for each other...'

Number of utterances: 3 (number of syllables) x 3 (types) x 3 (repetitions) = 27 (utterances)

Table 4: Morphophonological domains

Context	Morpho-phonological context	Morpho-phonological domains	Translation
1	SC + <u>verb stem</u>	[re] _{INFL} -[_{Stem} mêmá]...	'we invite...'
2	<u>SC</u> + verb stem	[o] _{INFL} -[_{Stem} lôya]...	'he bewitches...'
3	<u>SC</u> + T + verb stem	[o-tlo] _{INFL} -[_{Stem} lôya]...	'he will bewitch...'
4	SC + <u>I</u> + verb stem	[re-sa] _{INFL} -[_{Stem} lôya]...	'we still bewitch...'
5	SC + <u>OC</u> + verb stem	[re] _{INFL} -[_{MACRO} le-[_{Stem} lôya]]...	'we bewitch him...'

Number of utterances: 5 contexts x 3 (di-, tri-, quartossyllabic stems) x 3 (types) x 3 (repetitions)⁵ = 135 utterances

Subjects

Four speakers participated in the study, two male and two female. All participants were between 20 and 30 and grew up in the area around Polokwane which is predominantly Northern Sotho-speaking. They consider themselves Northern Sotho speakers and their speech shows indeed the characteristics of Northern Sotho, with common variation in the pronunciation of <fs> (see Ziervogel *et al.*, 1967: 122).

Procedure

Recordings were carried out in a quiet room at the Meraka Institute, Pretoria, and the University of the Witwatersrand, Johannesburg. The speech signal was received by a Sennheiser microphone and recorded onto cassette for the recordings in Pretoria. In Johannesburg, the signal was received by an M-Audio Microtrack digital recorder. The signal was digitised with a sampling frequency of 44100Hz and stored on a Macintosh laptop hard drive.

The target sentences were presented on a Macintosh laptop using PowerPoint software. The participants were seated in front of the laptop monitor. The microphone was placed by the side of the monitor, approximately 30cm away from the subject's lips. The session started with instructions displayed on the screen. The sentences were presented in Northern Sotho with an English translation (see Appendix). The English translation was provided to control for the tonal structure of the verbs, should ambiguities arise as tone is not indicated in the orthography. The participants were instructed to read out loud only the Northern Sotho sentences. The sentences were presented in random order together with sentences of another production study (see Zerbian & Barnard, unpublished data).

F0 extraction and measurements

The digitised data were analysed on a PC workstation. The individual target sentences were extracted and saved as separate .wav files. They were further analysed using the speech analysis software PRAAT (Boersma & Weenink, 2005). The data were first inspected visually (see Xu, 2001 for a corresponding procedure). For visual inspection, a PRAAT script (Xu, 1999) was run over the data that rendered time-normalised F0 curves for the relevant part of the target sentences. A four-syllable-window was chosen for which F0 was calculated, starting with the tone-bearing syllable (TBU). The segmentation of the syllables was performed manually by inspecting the speech waveform, while drawing on spectral and auditory information where desirable (Ladefoged, 1993;

Turk *et al.*, 2006). Time-normalisation for F0 was computed by taking a fixed number of points at equal time proportions (in this case ten). The target syllables were individually inspected for vocal pulses that were incorrectly calculated by PRAAT. Alterations were made manually. F0 values are missing for portions of segments that show no F0 (e.g. [h], [s], [t]). For a clearer graphical picture of F0-syllable alignment, the F0 curves of the four syllables were averaged across the three repetitions for each type.

Results

Length of the verb stem

A research question that influences the rest of our analysis relates to the influence of the length of the verb stem (i.e. the number of syllables of the verb stem) on high tone realisation, i.e. if a single high tone is realised differently across verb stems of different syllable length. To address this issue, we first present visual evidence, and then perform a statistical analysis.

Visual inspection

The verb types of context 1 are repeated in (8) from Table 3. Only the data from context 1 are presented visually here, however, inspection of the other contexts suggests a similar pattern (this can be checked by glancing at the other plots where data are depicted separately for different lengths of the verb stem).

- (8) Context 1
- | | | |
|-----------------------------|--------------|-----------------------------|
| [re] _{INFL} -[Stem | mêma]... | 'we invite...' |
| [re] _{INFL} -[Stem | rumula]... | 'we provoke...' |
| [re] _{INFL} -[Stem | rômêlana]... | 'we send for each other...' |

Figure 1 shows for each speaker individually the four-syllable-window in which the high tone originates on the verb stem initial syllable. Fundamental frequency (F0) in Hertz is displayed on the y-axis. For ease of comparison, the y-axis shows a pitch range of 120Hz across all diagrams, adjusted in absolute height to the individual's pitch range. The x-axis shows the four-syllable window for which F0 measurements have been taken: tick marks which are labelled indicate right syllable boundaries, whereas unlabelled tick marks represent the onset segment boundary within a syllable. Each display starts with the tone-bearing syllable. The solid line indicates disyllabic verb stems, the dashed line trisyllabic verb stems, and the dotted line quattrosyllabic verb stems.

The displays in Figure 1 show that across all speakers the pitch starts rising on the tone-bearing syllable from a relatively low level. The location of the high peak is subject to variation across speakers, occurring somewhere between the second and third syllable, counting from the tone-bearing syllable. From the pitch peak, the pitch trails off towards the end of the four-syllable window. The observation of interest for the current analysis is that verb stems of different syllable lengths seem quite similar in the way that pitch is realised (except possibly for speaker NM in context 1). To determine whether these differences are significant in this or other contexts, we carried out a statistical analysis.

Analysis

In order to ascertain the reliability of a significant variation across stems of different length, we conducted a statistical analysis on the data which is reported in the current section. In order to test for variation in the alignment of the pitch peak two series of two-factor ANOVA tests were conducted. In the first series, we investigated whether the pitch frequencies *around the pitch peak* were significantly different for different stem lengths, by treating the number of syllables in the stem and the syllable number (TBU or TBU+1) as independent variables, with the mean F0 value of the respective vowel as the dependent variable. For the second series, we similarly analysed pitch frequencies in the 'tail' following the peak, now taking as independent variables number of syllables in the stem and the syllable number (TBU+2 or TBU+3), and using the same independent variable.

Results are computed for each speaker and context individually. We refrained from pooling across speakers due to the idiosyncrasies of an individual's pitch range. Though we could have normalised

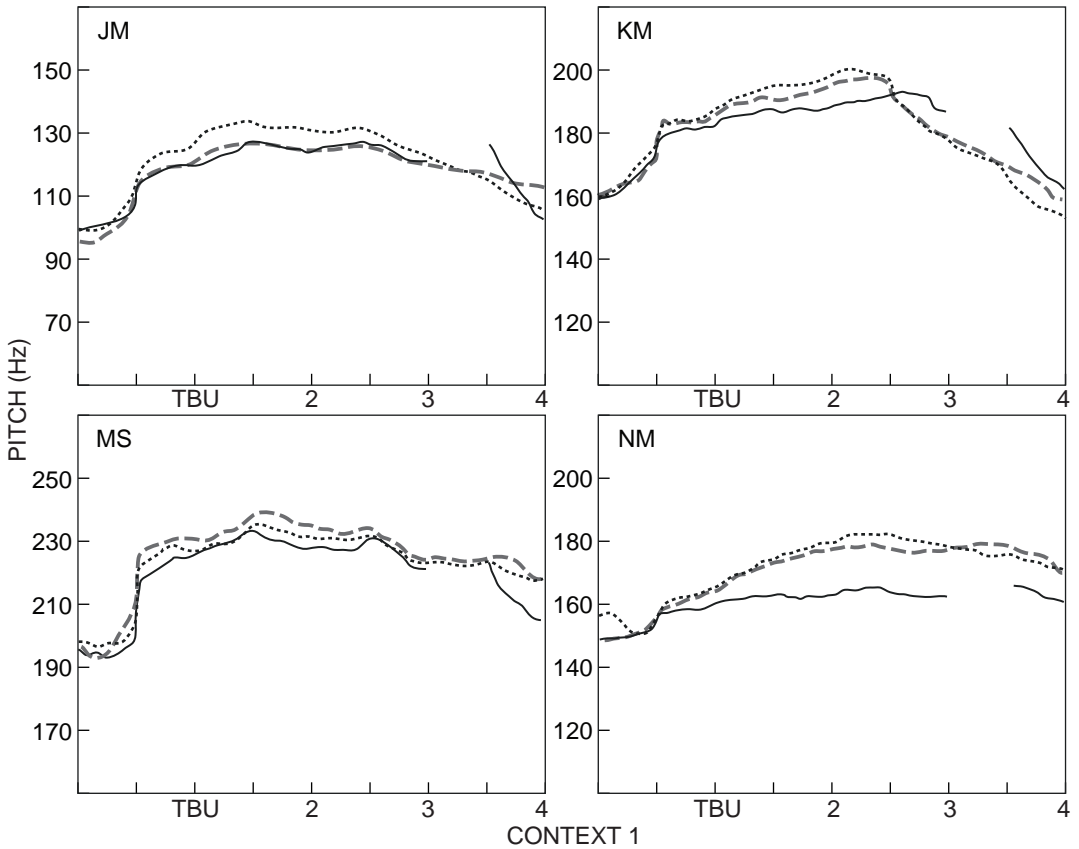


Figure 1: Realisation of a high tone across verb stems of different lengths in context 1

the pitch values across gender by converting them to semitones (see for example Xu, 1999 for a comparable procedure), this procedure does not capture the differences in pitch excursion that an individual uses in his/her speech.

Tables 3 and 4 summarise the results obtained for the peak and tail analyses, respectively. In these Tables, an X indicates that the number of syllables in the verb stem was found to be a significant factor in the ANOVA analysis (with $p < 0.05$ being used as the significance criterion).

These results show considerable differences between speakers, with two speakers (KM and NM) often showing significant differences in the production of both 'peak' and 'tail' syllables, whereas MS making hardly any distinction based on the number of syllables. JM displays some verb length differences in the tail syllables, but not in the peak. Thus, the number of syllables in the verb stem does influence the realisation of a single high tone significantly for some speakers and at least in some contexts. However, this influence is not consistent. It therefore seems as if factors such as segmental make-up interacting with speaker idiosyncrasies are the most likely explanation for the observed differences. Another factor in the observed differences could be pitch declination: as the number of syllables in the verb stems increases, the overall number of syllables in an utterance tends to increase. Due to declination across the utterance, the down slope of a high tone will be affected by the presence of additional syllables in the utterance. This topic clearly deserves more research, but it seems likely that the verbs of different lengths do not differ phonologically. We nevertheless analyse verb stems of different length separately in the analyses that follow.

Morphosyntactic constituency

Morphosyntactic constituency is the second parameter that is varied in a controlled fashion in the current study in order to investigate its influence on high tone realisation in Northern Sotho. Our data will be tested against two hypotheses: first, that the morphological nature of the constituent from which the high tone originates influences the realisation of a single high tone; and second, that an immediately following morphosyntactic boundary also influences high tone realisation.

All five contexts provided data that were used to test the hypotheses. The contexts are repeated in Table 7 from Table 4.

A comparison of the tone realisation in contexts 1, 2, 4, and 5 provides data concerning the first question, namely whether the morphological nature of the constituent from which the high tone originates has a determining influence of high tone realisation. A comparison of the tone realisation in contexts 2 and 3 provides data concerning the second question, namely whether an immediately following morphosyntactic boundary influences the high tone realisation. The current section addresses the first hypothesis only.

Visual inspection

The realisation of the high tone in contexts 1, 2, 4, and 5 is compared in order to decide if the morphological nature of the constituent from which the high tone originates influences high tone realisation. The data are repeated in (9) from Table 4. In context 1, the high tone originates on the initial syllable of a verb stem, (9a). In context 2, the high tone originates on the subject concord, (9b). In context 4, it originates on a tense marker, (9c). And finally in context 5 it originates on the object concord, (9d).

(9a) Context 1

[re] _{INFL} -[S _{stem} mêmá]...	'We invite...'
[re] _{INFL} -[S _{stem} rúmula]...	'We provoke...'
[re] _{INFL} -[S _{stem} rômêlana]...	'We send for each other...'

Table 5: ANOVA analysis of the average F0 of TBU and TBU+1

Speaker	JM	KM	MS	NM
Context 1		X		
Context 2		X		X
Context 3				X
Context 4		X		
Context 5			X	

Table 6: ANOVA analysis of the average F0 of TBU+2 and TBU+3

Speaker	JM	KM	MS	NM
Context 1		X		X
Context 2	X	X		X
Context 3	X			X
Context 4	X	X		
Context 5		X		

Table 7: Morpho-phonological domains

Context	Morpho-phonological context	Morpho-phonological domains	Translation
1	SC + verb stem	[re] _{INFL} -[S _{stem} mêmá]	'we invite'
2	SC + verb stem	[o] _{INFL} -[S _{stem} lôya]	'he bewitches'
3	SC + T + verb stem	[o-tlo] _{INFL} -[S _{stem} lôya]	'he will bewitch'

- (9b) Context 2
- | | |
|---|-----------------------------|
| [ɔ] _{INFL} -[nyenya] _{Stem} ... | 'S/he despises...' |
| [ɔ] _{INFL} -[mênola] _{Stem} ... | 'S/he turns upside down...' |
| [ɔ] _{INFL} -[mamarêla] _{Stem} ... | 'S/he sticks to...' |
- (9c) Context 4
- | | |
|---|--------------------------------|
| [re-sa] _{INFL} -[lwêla] _{Stem} ... | 'We still fight for...' |
| [re-sa] _{INFL} -[mênola] _{Stem} ... | 'We still turn upside down...' |
| [re-sa] _{INFL} -[lamolêla] _{Stem} ... | 'We still rescue for...' |
- (9d) Context 5
- | | |
|--|-----------------------------|
| [re]-[_{Macro} le-] _{Stem} nyenya]]... | 'We despise him...' |
| [re]-[_{Macro} le-] _{Stem} mênola]]... | 'We turn it upside down...' |
| [re]-[_{Macro} le-] _{Stem} mamarêla]]... | 'We stick to it...' |

Figure 2 depicts the pitch contours of these contexts individually for all speakers. The contexts are displayed in columns starting at the left with context 1. (The solid line indicates disyllabic verb stems, the dashed line trisyllabic verb stems, and the dotted line quattrosyllabic verb stems)

There are definite intra-speaker differences but the overall trends depicted in the displays can be described as follows: As mentioned above for context 1, the pitch starts rising on the tone-bearing syllable from a relatively low level (which corresponds to the preceding low tone if present). The location of the high peak is subject to variation across speakers, occurring somewhere between the second and third syllable, counting from the tone-bearing syllable. The pitch stays high for a while, resulting in a plateau effect. In context 2, the pitch starts on a relatively high level and stays up throughout the second syllable before it starts to decline slowly. In context 4, we can make a similar observation. After the voiceless onset of the tone-bearing syllable, F0 starts out relatively high. It seems to start declining slightly earlier than in the previous context, namely before the end of the second syllable already for at least three out of the four speakers. Finally, in context 5, we find the pitch starting out from a relatively low level, just as in context 1. The rise to the maximum peak is delayed and only occurs in the third syllable.

Analysis

To test the systematicity of the observed variations, we conducted a three-factor ANOVA test. The mean F0 values in the vowel of interest is taken as the dependent variable, and the three independent variables are (a) the context from which the vowels originated, (b) the position of the syllable in the utterance (TBU or TBU+1), and (c) the number of syllables in the verb stem. Table 8 summarises the computed *p*-values for the context factor (with one degree of freedom), when contexts 1 and 4 are compared.

As the results show, the realisation of the high tone does not differ significantly in contexts 1 and 4 for any of the speakers. As for the extension of the domain of high tone spread to the right, see the section Visual Inspection below. When comparing contexts 2 and 3 in the same respect, the comparison is significantly different for only one of the four speakers, as shown in Table 9 (*p*-values are again for the context factor, with one degree of freedom).

It needs to be taken into consideration that contexts 2 and 3 both have an onsetless syllable in utterance-initial position. As a consequence, slight F0 variations are not surprising as the vocal folds only begin to vibrate and might thus show some irregularities as the manual inspection confirmed.

We are able to compare context pairs 1–4 and 2–3, respectively, as these pairs have similar syllable structures. However, a direct comparison of other pairs such as 1–2 or 1–3 is not sensible: in those pairs there is a critical segmental difference in addition to different morphosyntactic origin of the high tone. Due to the structure of the language, this segmental difference could not be avoided. Whereas the high tone from the verb stem originates on the syllable with a sonorant onset, the high tones from the subject concord (context 2, 3) and the mood marker *sá* (context 4) originate from a syllable that is either onsetless in utterance-initial position (as in the case of the subject concord *ó*) or has a voiceless onset (as in the case of *sá*). We will argue in detail in the final section why we think that this difference in segmental syllable make-up is responsible for the difference

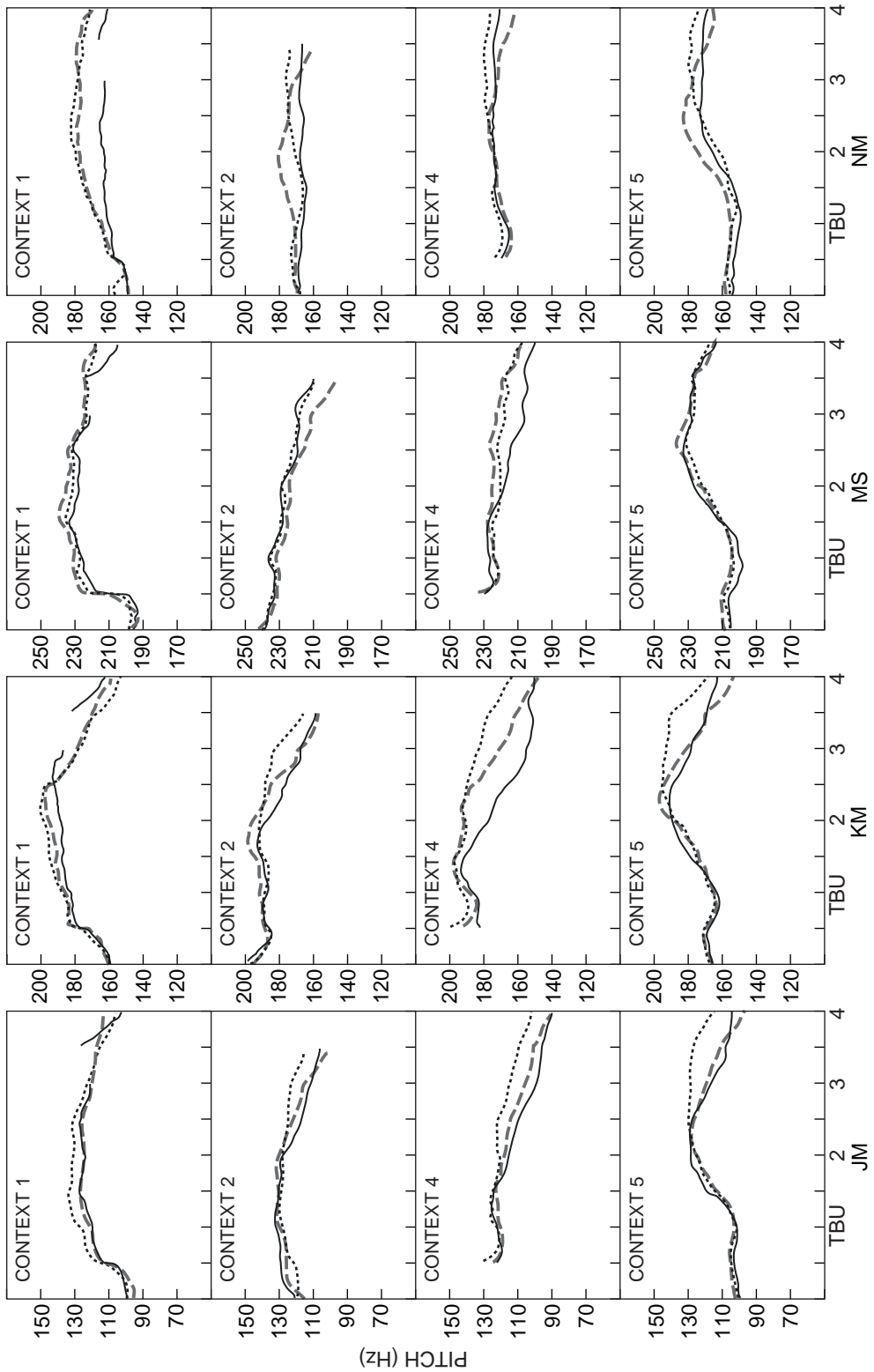


Figure 2: Realisation of a high tone across verb stems of different lengths in contexts 1, 2, 4 and 5

in the realisation of the high peak. Whereas we find a peak delay in context 1 in which the peak starts rising throughout the sonorant onset, context 2 and 4 lack the initial low pitch value due to the absence of a sonorant onset (being utterance-initial without an onset in context 2, and with a voiceless onset in context 4). We discuss this peak delay in the final section.

When comparing the last of the contexts, namely context 5, to the other contexts we find consistent results across speakers (again, *p*-values are for the context factor, with one degree of freedom). The statistical analysis confirms the impression gained from visual inspection of the data, namely that there is a significant difference in high tone realisation between context 5 and all other contexts. The difference is very clear and uniformly shared by all speakers.

To summarise the results concerning the realisation of high tones at the left edge of the domain across different morphosyntactic contexts: we find statistically significant differences with two (groups of) contexts. First, with context 1 we find a slight peak delay when compared to contexts 2 and 3. Although the difference is consistent, we have reason to believe that it arises from phonetic causes and is not a systematic phonological difference. The reason why we think we are dealing with a phonetic difference rather than a phonological one is because the difference can be explained by the differences in the segmental environment between these contexts. This will be addressed in more detail in the section ‘Peak delay – phonetic or phonological?’.

Second, we find a clear peak shift in context 5 which cannot be explained by reference to the segmental environment and which must thus be phonological.

High tones across word-internal morphological boundaries

Visual inspection

The pitch contours from contexts 2 and 3 are compared to see if an immediately following word-internal morphological boundary influences the realisation of a high tone. The analysis in Table 9 has already shown that there is probably no systematic difference at the left edge of the

Table 8: ANOVA for average F0 of TBU and TBU+1 and context when comparing contexts 1 and 4

Speaker	p-value	Significant?
JM	0.091459	
KM	0.211576	
MS	0.065220	
NM	0.116247	

Table 9: ANOVA for average F0 of TBU and TBU+1 and context when comparing contexts 2 and 3

Speaker	p-value	Significant?
JM	0.004237	X
KM	0.134716	
MS	0.084917	
NM	0.210860	

Table 10: ANOVA for average F0 of TBU and TBU+1 and context when comparing contexts 1 and 5

Speaker	p-value	Significant?
JM	0.000056	X
KM	0.000000	X
MS	0.000129	X
NM	0.000449	X

peak alignment. However, work on neighbouring Bantu languages has suggested that we can expect a difference at the right edge of the high tone domain.

One set of target sentences shows the high-toned subject concord, followed by a low-toned verb stem (representing present tense), (10a), whereas the other shows the high-toned subject concord marker, followed by a low-toned tense marker (indicating future) and a low-toned verb stem, (10b). Thus, the difference between the two lies solely in the presence of the low-toned future marker. Morphologically, this implies that in context 2, (10a), the high-toned subject marker is immediately followed by the stem boundary, whereas the tense marker intervenes between the high-toned syllable and the stem boundary in context 3, (10b).

Contexts 2 (10a) and 3 (10b)

(10a)	[o] _{INFL} -[nyenya] _{Stem} ...	'S/he despises...'
	[o] _{INFL} -[mênola] _{Stem} ...	'S/he turns [sth.] upside down...'
	[o] _{INFL} -[mamarêla] _{Stem} ...	'S/he sticks to...'
(10b)	[o-tlo] _{INFL} -[nyenya] _{Stem} ...	'S/he will despise...'
	[o-tlo] _{INFL} -[mênola] _{Stem} ...	'S/he will turn [sth.] upside down...'
	[o-tlo] _{INFL} -[mamarêla] _{Stem} ...	'S/he will stick to...'

Figure 3 compares the contours of the different sets graphically. The displays in the left column show context 2 for all four speakers, the displays in the right column show context 3. Again, the solid line indicates disyllabic verb stems, the dashed line trisyllabic verb stems, and the dotted line quattrosyllabic verb stems.

The target structures in (10) differ from each other segmentally, in that only in (10b) the first syllable after the high-toned syllable starts with a voiceless stop which is released laterally (<tl> = [tʰ]). This difference is of course reflected in the pitch contour in that there are no F0 measurements in the corresponding section.

Despite this difference, we see a similar pitch contour when comparing the two contexts across speakers: the pitch starts high on the tone-bearing syllable. From this high starting point the pitch slowly drops over the course of the utterance. We find some microprosodic disturbances in F0. These are due to starting vocal fold vibrations in utterance-initial position (both in contexts 2 and 3) and to a preceding voiceless syllable onset [tʰ] in context 3. As known from the literature on other languages, voiceless obstruents influence the pitch of the following vowel by slightly raising it (e.g. Hombert, 1978).

Analysis

We again conducted a three-factor ANOVA test on the last two vowels of our four syllable window in order to see if the two contexts differ in the average F0 values. Table 11 presents the results for the context factor. Again, *p*-values are for the context factor, with one degree of freedom. There is no significant difference in the realisation of the vowels in syllable 3 and 4 starting from the tone-bearing unit.

Relating back to the initial hypothesis, the results in Table 11, together with those in Table 9, suggest that an immediately following stem boundary does not influence the realisation of high tones originating on a high-toned subject concord. This finding will be discussed further in the section 'Implications for Sotho tonology'.

Discussion

The results of this study are relevant for the tonal description of the Northern Sotho language and phonological accounts of its tonology with respect to the numbers of syllables in the verb stem, the morphosyntactic origin of the high tone and the importance of word-internal morphological boundaries for high tone realisation, as pointed out in the Introduction and second section. However, the results also exemplify more general phonetic properties in the tonal grammar of Northern Sotho, such as peak delay and the influence of segmental structure on tone realisation. Knowledge of the latter informs the interpretation of and generalisations drawn from acoustic data. Thus, the discussion starts out with re-addressing the latter issue before the implications for Northern Sotho tonology are summarised.

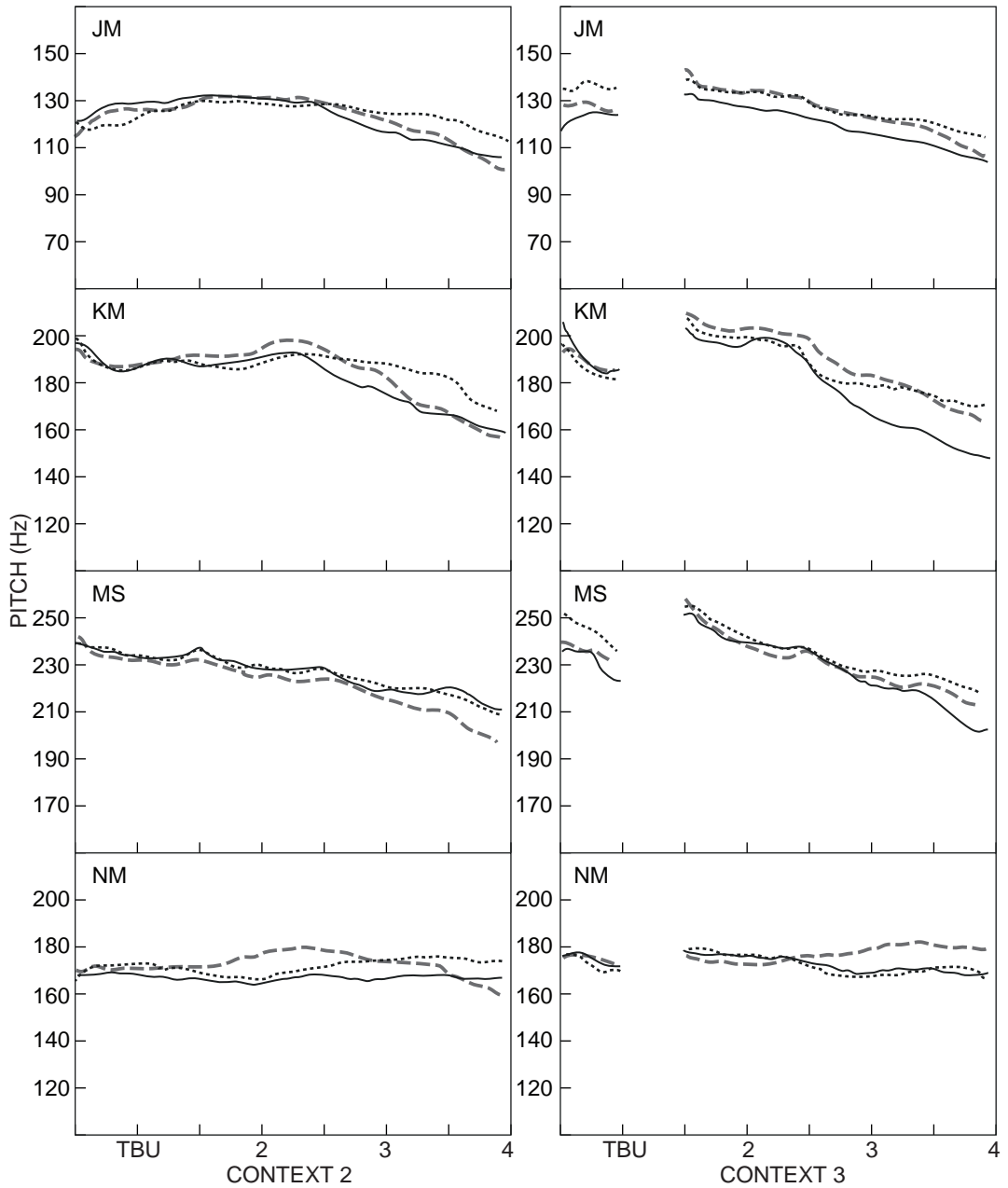


Figure 3: Realisation of high tone across different contexts

Peak delay – phonetic or phonological?

In investigating high tone alignment in Chichewa, an Eastern Bantu language, Myers (1999) finds that the pitch peak associated with a high tone is actually not reached in the syllable with which the high tone is associated underlyingly. In a controlled production experiment he investigates the alignment of high tones in the speech of three speakers. He defines a high tone phonetically as

Table 11: ANOVA for average F0 of TBU+2 and TBU+3 and context when comparing contexts 2 and 3

Speaker	p-value	Significant?
JM	0.699352	
KM	0.084906	
MS	0.479678	
NM	0.661118	

a local peak in F0 and finds that F0 starts to rise near the beginning of the syllable that the high tone originates from. However, the pitch peak is only reached at the end of that syllable or at the beginning of the next. This description matches the observation made in this study in context 1.

This delay in reaching the peak maximum only in the beginning of the subsequent syllable is a cross-linguistically attested phenomenon and known as peak delay. This peak delay is also known from studies of tone alignment in better-investigated languages, such as European languages or Mandarin Chinese (Silverman & Pierrehumbert, 1990; Prieto *et al.*, 1995; Xu, 2001). We thus think it is safe to conclude that we find a similar peak delay in Northern Sotho tonology on syllables that contain a specific segmental make-up.

From a review of the literature we know that Northern Sotho has been claimed to exhibit a phonological process called high tone spread on to (at least) the immediately following syllable. Phonological high tone spread would phonetically involve a separate high tone target on vowels that follow the high tone-bearing vowel. Myers (1999) has equated the peak delay observed in Chichewa with the high tone spread reported in the phonological literature of this language (e.g. Kanerva, 1990). Thus, the process is considered to be phonetic in this language. According to this phonetic view, the pitch peak is realised later in the syllable as the articulators need a specific time to reach their target settings (see also Xu, 2009). As this time is longer than the duration of short syllables in Chichewa, the pitch peak is only reached in the following syllable. The duration of long syllables, however, exceeds the obligatory articulator setting and thus the pitch peak is reached within the syllable.

There is some evidence against applying this purely phonetic view to all cases of bound high tone spread. Firstly, Myers' study itself provides contradictory evidence. He reports that the peak delay remains stable under changes in speech tempo (Myers, 1999: 219). Stability across different speech tempi is normally a sign for a phonological rather than phonetic effect (Yip, 2002: 10) and as such contradicts a purely phonetic account. Secondly, a previous study by Zerbian (2007) suggests that high tone spread is absent across word boundaries in Northern Sotho. If high tone spread was merely a phonetic effect, due to physiological constraints, it should not be influenced by word boundaries.

The data collected in the present study is not able to conclusively answer the question whether in Northern Sotho there is phonological high tone spread or rather a phonetic carry-over effect with high tones. The data that concern the right edge of the high tone domain are rather variable across subjects and contexts and deserve more detailed investigation before any conclusive generalisation can be made. In general a slow fading out of the pitch can be observed. We assume that this is due to a lacking following pitch target, similar to the neutral tone in Chinese (Chen & Xu, 2006) which is implemented slowly and ineffectively. This assumption has of course to be explored in further research. The data presented here, however, bring to light two parameters of variation that will need to be taken into account when further investigating this question for Northern Sotho. These parameters are segmental make-up of the tone-bearing syllable and its position within an utterance.

In order to realise a hightone, the pitch target is aimed at from around the onset of the tone-bearing syllable. If the tone-bearing syllable shows a voiced onset, the pitch raises throughout the all-voiced syllable until the peak is reached in the immediately following syllable (or even the syllable thereafter). The velocity of pitch rise might be determined by physiological constraints of the vocal folds. But this is not necessarily the only possibility. Alternatively, a rise in pitch might

be sufficient for the hearer to perceive a high tone. Perception experiments will have to test this hypothesis. If the tone-bearing syllable starts on a voiceless onset, however, the vocal folds can adjust their tension during the voiceless period to immediately start out on a higher fundamental frequency when voicing sets in again on the syllable nucleus. Similarly, if the tone-bearing syllable constitutes the first syllable of an utterance, vocal fold tension can be adjusted before the onset of the utterance in order to produce a pitch peak on the tone-bearing syllable itself. This would explain the less steep pitch rise in the tone-bearing syllables of context 2 and 4. This phenomenon is known in the literature since at least Bruce (1977) as tone truncation.

The hypotheses in the preceding paragraph make the predictions that in word-medial position tone alignment changes with the voicing of the onset of the tone-bearing syllable. This prediction needs to be tested systematically in order to further support this claim. The data collected for the present study were deliberately designed to only feature sonorants in onset position. As such, they are not appropriate to decide this issue. However, it turned out that <h> is realised as a voiceless glottal fricative in the verbs <humana> and <humanêla> in the speech of the participants of the study (see Ziervogel *et al.*, 1967: 126). Consequently, also with trisyllabic and quatosyllabic verb stems we find one verb type featuring a voiceless syllable onset in utterance-medial position.

Figure 4 shows the alignment of the high tone in context 1 (i.e. from a verb stem initial syllable). The solid line shows the average of three trisyllabic verbs that start with a voiceless syllable onset, and the dotted line shows the average for the remaining six tokens of the other two verb stems

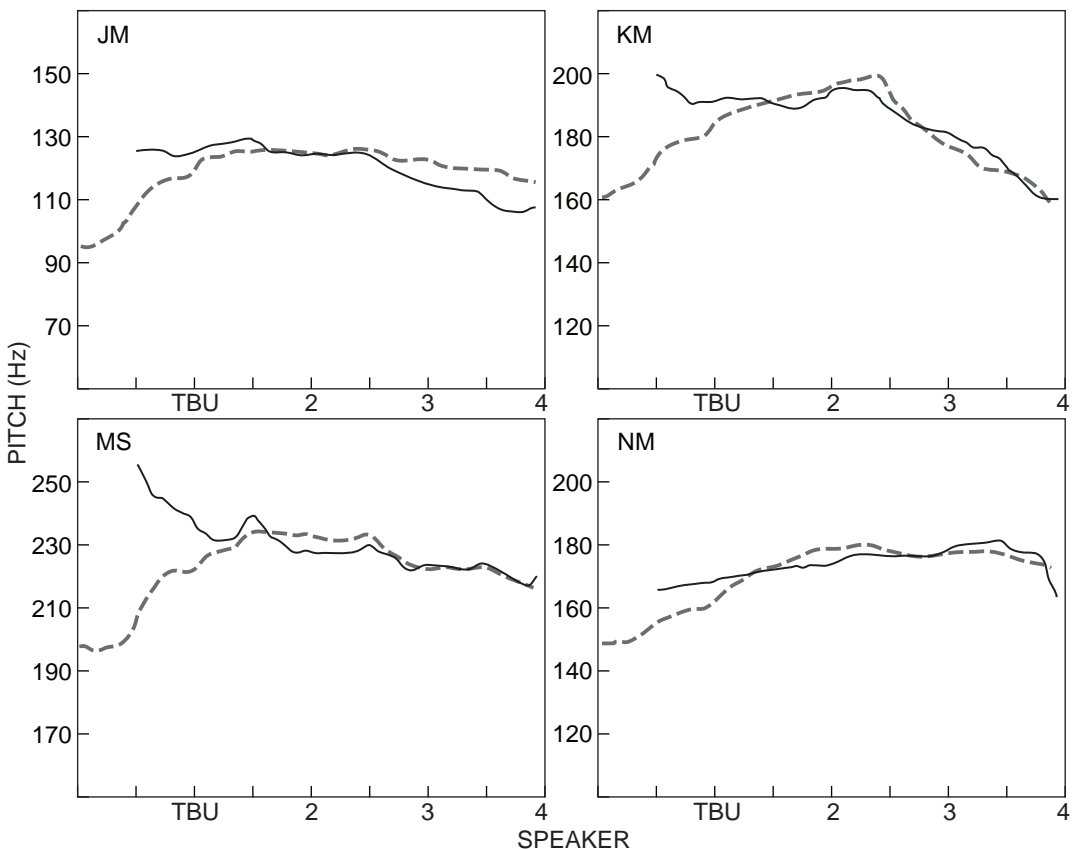


Figure 4: Pitch depending on the onset segments

that start with a sonorant onset. We see that across all speakers the pitch contours differ in these two conditions. Whereas we find the described peak delay with sonorant onsets, we find the pitch starting at a high level when the verb stem in question starts on a voiceless segment.

This finding is not surprising as the pitch raising effect of the voiceless obstruent and the high tone target of the syllable conspire. Figure 4 thus supports our claim that the segmental syllable make-up (as well as position within the utterance) need to be taken into consideration when making claims about the phonology or phonetics of high tone spread and peak delay, also in Bantu languages.

To summarise, phonologically, a high-toned object concord causes a peak shift one syllable to the right. Reference to morpho-phonology is needed in order to account for the observed high tone shift which occurs with a high tone originating on the object concord in Northern Sotho.

Phonetically, our study confirms the occurrence of peak delay also in Northern Sotho, i.e. the observation that the F₀ peak associated with a high tone is not necessarily reached in the syllable it is associated with. Thus, *peak delay* described in phonetic studies could correlate with what has been described as *high tone spread* in the phonological literature. Our results support the view that the phonological process of high tone spread can actually be captured phonetically in Northern Sotho in the contexts investigated here. This result aligns with an identical claim made by Myers (1999) for Chichewa.

Furthermore, the segmental make-up of the tone-bearing syllable leads to a systematic surface variation in tone alignment. Our data show that utterance-initial onsetless syllables or utterance-medial syllables with a voiceless onset show slightly different F₀ contours which differ only phonetically, though.

Implications for Sotho tonology

The current study also has direct implications for the description and account of tone in Northern Sotho. The underlying tonal specification of all morphemes included in the study was largely as expected given the tone-marked dictionary (Ziervogel & Mokgokong, 1979) and a grammar that comprises some tone marking as well (Ziervogel *et al.*, 1969). However, the peak shift inducing object concord has not been reported previously in the literature on Northern Sotho. Interestingly, the same phenomenon has been reported for the eastern neighbour Tsonga (Kisseberth & Odden, 2003) and has been confirmed for the southern neighbour Southern Sotho as well (Moloi and Khoali, personal communication), though no detailed data and account is available.

In the second section of the article we alluded to the difficulties in interpreting tonal data available in the literature due to a variety of differences in the data presented, among them varying number of syllables in the verb stem, differences in the segmental make-up of the tone-bearing syllables, different transcribers and different dialects.

We think that the present article has controlled these aspects or has at least delineated the parameters. It remains to suggest a tonal transcription for the structures under consideration and to define how we used the tone labels. Table 12 is a repetition of Table 2, to which we added the data of our own study in the last column. We use an accent (indicating high tone) on a syllable that has an underlying high tone target and which is started to be implemented on that tone-bearing syllable. If this high tone target is only reached in the following syllable due to segmental constraints (as for example in context 1), this will not be reflected in our transcription. If, however, the high tone target is not started to be implemented on the tone-bearing syllable, as in context 5, we leave this syllable without an accent. It will nevertheless be underlined because it contributes the high tone. Only the following syllable – on which the high tone target is actually started to be implemented by a rising pitch – will receive an acute accent.

Regarding the post-peak pitch contour, dependent on the speaker and the number of syllables, the pitch peak either occurs in the third syllable (speakers KM and NM) or extends in a high plateau into the third syllable (speaker JM, and to a lesser degree speaker MS). The statistical analyses revealed quite some variation depending on speakers and contexts. Intuitively, we think that the post-peak contour slowly fades out as a result of the fact that no further high tone target follows, similar to the neutral tone in Chinese (Chen & Xu, 2006) which is implemented slowly and ineffectively. As mentioned above, further evidence is needed to substantiate this view. As for

Table 12: Tone patterns of verb forms in Sotho-Tswana varieties

	Northern Sotho	Sepedi	Sesotho	Sesotho	Tswana	Northern Sotho
<i>Structure</i>	<i>Ziervogel et al.</i>	<i>Lombard</i>	<i>Khoali</i>	<i>Lets'eng</i>	<i>Chebanne et al.</i>	<i>Own study</i>
1 SC + <u>verb stem</u>	re-rútá re-rútísa (p. 134)	ke-kólóbétša kgóši (p. 83)	ke-rékísa nama (p. 239) o-kgúrúmetša motho (p. 190)	ke-bótsá tsila (p. 22) ke-kótúla mabelé (p.22)	ke-thúsa ngáka lo-tshámíkisa metsí (p. 92)	re-méma re rémelela
2 <u>SC</u> + verb stem	ó-réta	bá-hlágolela malóme lé-tlíša dipúdi (p. 84)	ó-bátla nama (p. 205) ó-kgárametša motho (p. 190)	bá-bítsana ká majwe (p. 26)	ó-bátla gó-tsamaya ó-ápára sesádi (p. 94f)	ó-loya
3 <u>SC</u> + T + verb stem	ó-tló-reta	á-tló-epolla (p. 157)	ó-tlá-tsamaaya (p. 224)	bá-tlá-bitsana ká majwe (p. 72)	bá-tláá-batla ó- tláá-tselela (p. 117f)	ó-tlo-loya
4 SC + <u>I</u> + verb stem	re-sá-reta	le-sá-epolla re-sá-apara	not available	ke-sá-bula lematí ke-sá-lokisa kolói (p. 81)	not available	re-sá-loya
5 SC + <u>OC</u> + verb stem	re-é-reta re-é-rétela	go-dí-bápala (p. 58)	not available	o-mg-étela mokibelo (p. 25)	ke-ló-bálá naé ke-bá-báléla lokwálo (p. 93f)	re-le-lóya

tone-marking in these instances, we do not think that any syllable after the tone-bearing syllable really has an additional high tone target. In all instances, it seems that the pitch starts dropping after the peak. The slope of the pitch drop might depend on the speaker and on the overall number of syllables in the utterance (as this will affect declination).

What our study did bring to light, however, is that an immediately intervening word-internal boundary does not affect the realisation of a high tone, contrary to what Myers has claimed for Shona (1987). Thus, contexts 2 and 3 do not differ in any systematic way in the domain of high tone spread. Neither is context 4 different from contexts 1–3 tonally.

Furthermore, the role of number of syllables in the verb stem remains inconclusive. We did find some variation between speakers. So we can neither confirm an identical behaviour in high tone realisation of stems of different lengths as in most publications on Sotho tone, nor can we confirm the dichotomy reported in Monareng's work (1992) for the Setswapo dialect of Northern Sotho where high tones are reported to spread differently in disyllabic and polysyllabic verbs.

Notes

- 1 It has been noted that next to idiosyncratic differences there are also age-related differences in tonal realisation. Lombard (1976: 2) cites Jones/Plaatje (1916) that 'The younger generation of Bechuana [Setswana speakers, SZ] are to some extent losing the original Sechuana [Setswana] tones'.
- 2 One might argue that there are moods which are expressed solely by tonal patterns. Whereas that might be the case in other Bantu languages, it is not clear that this case indeed exists in Sotho-Tswana: often the subject marker differs in its underlying tone and/or shape, so that we have to assume a different morpheme.
- 3 In the presentation, syllables are underlined when indicating the tone-bearing unit. As Northern Sotho does not have contrastive vowel length, the syllable and the vowel are inseparable as tone-bearing units. As (most of) the verbal prefixes are monosyllabic, morpheme and tone-bearing unit fall together in the examples as well.
- 4 Thus, only the context with a high-toned OC was tested and not the possible structures below:

SC + OC + verb stem
 SC + TMA + OC + verb stem
 SC + TMA + OC + verb stem

- ⁵ One trisyllabic, low-toned verb (*laola*) had to be excluded with all its repetitions as it contained an onsetless medial syllable that would not allow alignment with the other verbs in the time-normalised display. Thus, only 132 utterances (instead of 135) were analysed for each speaker.

Acknowledgements — We want to thank audiences at the 9th LASU conference (Maputo), ZAS (Berlin) as well as Yi Xu, Laura J Downing and three anonymous reviewers for helpful feedback. We gratefully acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG; grant to the ZAS) as well as the Council for Scientific and Industrial Research (CSIR) in carrying out this research.

References

- Barrett-Keach C.** 1986. Word-internal evidence from Swahili for Aux/INFL. *Linguistic Inquiry* 17: 559–564.
- Boersma P & Weenink D.** 2005. Praat: doing phonetics by computer. Computer software. Available at www.fon.hum.uva.nl/praat/ [accessed 25 October 2009].
- Bruce G.** 1977. *Swedish word accents in sentence perspective*. Travaux de l'Institut de Linguistique de Lund 12. Lund: CWK Gleerup.
- Cassimjee F & Kisseberth CW.** 1998. Optimality domains theory and Bantu tonology: a case study from Isixhosa and Shingazidja. In Hyman LM & Kisseberth CW (eds) *Aspects of Bantu tone*. Stanford: CSLI Publications, pp. 33–132.
- Chebanne AM, Creissels D & Nkhwa HW.** 1997. *Tonal morphology of the Setswana verb*. Munich: LINCOM Europe.
- Chen Y & Xu Y.** 2006. Production of weak elements in speech – evidence from F0 patterns of neutral tone in standard Chinese. *Phonetica* 63: 47–75.
- Cole DT & Mokaila DM.** 1962. *A course in Tswana*. Washington: Georgetown University Press.
- Cole DT.** In press. Setswana & English illustrated dictionary.
- Downing LJ.** 1999. Prosodic stem ≠ prosodic word in Bantu. In Hall TA & Kleinhenz U (eds) *Studies on the phonological word*. Amsterdam: John Benjamins, pp. 73–98.
- Downing LJ.** 2001. How ambiguity of analysis motivates stem tone change in Durban Zulu. *UBC Working papers in Linguistics* 4: 39–55.
- Hombert J-M.** 1978. Consonant types, vowel quality, and tone. In Fromkin VA (ed.) *Tone: a linguistic survey*. New York: Academic Press, pp. 77–111.
- Hyman LM.** 1993. Conceptual issues in the comparative study of the Bantu verb stem. In Mufwene SS & Moshi L (eds) *Topics in African linguistics*. Amsterdam: John Benjamins, pp. 3–34.
- Hyman LM & Mathangwane JT.** 1998. Tonal domains and depressor consonants in Ikalanga. In Hyman LM & Kisseberth CW (eds) *Theoretical aspects of Bantu tone*. Stanford: CSLI, pp. 195–229.
- Hyman LM & Ngunga A.** 1994. On the non-universality of tonal association 'conventions'. *Phonology* 11: 25–68.
- Kanerva JM.** 1990. *Focus and phrasing in Chichewa phonology*. New York, London: Garland Publishing.
- Khoali BT.** 1991. A Sesotho tonal grammar. PhD thesis, University of Illinois at Urbana-Champaign.
- Kisseberth CW & Odden D.** 2003. Tone. In Nurse D & Philippson G (eds) *The Bantu languages*. London, New York: Routledge, pp. 59–70.
- Ladefoged P.** 1993. *A course in phonetics*. Boston: Thomson Wadsworth.
- Lets'eng MC.** 1995. Les tiroirs verbaux du Sesotho. Memoire, University Stendhal (Grenoble III).
- Lombard DP.** 1976. Aspekte van toon in Noord-Sotho. PhD thesis, University of South Africa.
- Mchombo SA.** 1993. Reflexive and reciprocal in Chichewa. In Mchombo SA (ed.) *Theoretical aspects of Bantu grammar*. Stanford: CSLI, pp. 181–207.
- Mmusi SO.** 1992. Obligatory contour principle effects and violations: the case of Setswana verbal

- tone. PhD thesis, University of Illinois at Urbana-Champaign.
- Monareng WM.** 1992. A domain-based approach to Northern Sotho tonology: a Setswapo dialect. PhD thesis, University of Illinois at Urbana-Champaign.
- Mutaka N.** 1994. *The lexical tonology of Kinande*. Munich: Lincom Europe.
- Myers S.** 1987. Tone and the structure of words in Shona. PhD thesis, University of Massachusetts.
- Myers S.** 1998a. AUX in Bantu morphology and phonology. In Hyman LM & Kisseberth CW (eds) *Theoretical aspects of Bantu tone*. Stanford: CSLI, pp. 231–264.
- Myers S.** 1998b. Surface underspecification of tone in Chichewa. *Phonology* **15**: 367–391.
- Myers S.** 1999. Tone association and F0 timing in Chichewa. *Studies in African Linguistics* **28(2)**: 215–239.
- Myers S.** 2003. F0 timing in Kinyarwanda. *Phonetica* **60**: 71–97.
- Prieto P, van Santen J & Hirschberg J.** 1995. Tonal alignment patterns in Spanish. *Journal of Phonetics* **23**: 429–451.
- Silverman K & Pierrehumbert J.** 1990. The timing of prenuclear high accents in English. In Kingston J & Beckman M (eds) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press, pp. 72–106.
- Turk A, Nakai S & Sugahara M.** 2006. Acoustic segment durations in prosodic research: a practical guide. In Sudhoff S, Lenertová D, Meyer R, Pappert S, Augurzyk P, Mleinek I, Richter N & Schliesser J (eds) *Methods in empirical prosody research*. Berlin: Walter de Gruyter, pp. 1–28.
- Xu Y.** 1999. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* **27**: 55–105.
- Xu Y.** 2001. Fundamental frequency peak delay in Mandarin. *Phonetica* **58**: 26–52.
- Xu Y.** 2009. Timing and coordination in tone and intonation – an articulatory-functional perspective. *Lingua* **119(6)**: 906–927.
- Xu Y & Xu CX.** 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* **33**: 159–197.
- Yip M.** 2002. *Tone*. Cambridge: Cambridge University Press.
- Zerbian S.** 2006. The expression of information structure in Northern Sotho. PhD thesis, Humboldt University Berlin.
- Zerbian S.** 2007. Phonological phrasing in Northern Sotho. *The Linguistic Review* **24**: 233–262.
- Zerbian S & Barnard E.** The realization of two high tones in Northern Sotho. University of the Witwatersrand & Meraka Institute. Unpublished manuscript.
- Ziervogel D, Lombard DP & Mokgokong PC.** 1969. *A handbook of the Northern Sotho language*. Pretoria: Schaik.
- Ziervogel D, Louw JA, Ferreira JA, Baumbach EJM & DP Lombard.** 1967. *Handbook of the speech sounds and sound changes of the Bantu languages of South Africa*. Pretoria: University of South Africa.
- Ziervogel D & Mokgokong PC.** 1979. *Klein Noord-Sotho woordeboek*. Pretoria: van Schaik.

Appendix

Sentences are given in standard orthography as they were presented in the study. Underlying high tones are marked only for the presentation in this article.

Disyllabic verb stems

- | | | |
|----------|---|---|
| (2.1) a. | Re- <u>h</u> ula mokgalabje | 'We are robbing the old man' |
| b. | Re- <u>m</u> ema mokgalabje | 'We are inviting the old man' |
| c. | Re- <u>r</u> oma lesogana | 'We are sending the young man' |
| (2.2) a. | <u>Q</u> -loya lesogana | 'He bewitches the young man' |
| b. | <u>Q</u> -lwela mokgalabje | 'He is fighting for the old man' |
| c. | <u>Q</u> -nyenya lesogana | 'He despises the young man' |
| (2.3) a. | <u>Q</u> -tlo-loya lesogana | 'He will bewitch the young man' |
| b. | <u>Q</u> -tlo-lwela mokgalabje | 'He will fight for the old man' |
| c. | <u>Q</u> -tlo-nyenya lesogana | 'He will despise the young man' |
| (2.4) a. | Re- <u>s</u> a-loya lesogana | 'We are still bewitching the young man' |
| b. | Re- <u>s</u> a-lwela mokgalabje | 'We are still fighting for the old man' |
| c. | Re- <u>s</u> a-nyenya lesogana | 'We are still despising the young man' |
| (2.5) a. | Re- <u>l</u> e-loya lefsifsing (lephodisa) | 'We are bewitching him in the dark (the policeman)' |
| b. | Re- <u>l</u> e-lwela lefsifsing (lesea) | 'We are fighting for him in the dark (the baby)' |
| c. | Re- <u>l</u> e-nyenya lefsifsing (lesogana) | 'We are despising him in the dark (the man)' |

Trisyllabic verb stems

- | | | |
|----------|--|--|
| (3.1) a. | Re- <u>h</u> umana mokgalabje | 'We find the old man' |
| b. | Re- <u>r</u> umula lesogana | 'We provoke the young man' |
| c. | Re- <u>r</u> emela mokgalabje | 'We are chopping for the old man' |
| (3.2) a. | <u>Q</u> -namola mokgalabje | 'He is rescuing the old man' |
| b. | <u>Q</u> -menola leswika | 'He is turning the stone upside down' |
| (3.3) a. | <u>Q</u> -tlo-namola mokgalabje | 'He will rescue the old man' |
| b. | <u>Q</u> -tlo-menola leswika | 'He will turn the stone upside down' |
| (3.4) a. | Re- <u>s</u> a-namola mokgalabje | 'We are still rescuing the old man' |
| b. | Re- <u>s</u> a-menola leswika | 'We are still turning the stone upside down' |
| (3.5) a. | Re- <u>l</u> e-namola lefsifsing (lesea) | 'We rescue it in the dark (the baby)' |
| b. | Re- <u>l</u> e-menola lefsifsing (leswika) | 'We turn it upside down in the dark (the stone)' |

Quatrosyllabic verb stems

- | | | |
|----------|--|---|
| (4.1) a. | Re- <u>h</u> umanela lesogana | 'We are finding for the young man' |
| b. | Re- <u>r</u> umulela mokgalabje | 'We are provoking for the old man' |
| c. | Re- <u>r</u> omelana lefsifsing | 'We are sending for each other in the dark' |
| (4.2) a. | <u>Q</u> -lamolela mokgalabje | 'He is rescuing for the old man' |
| b. | <u>Q</u> -mamarela lesogana | 'He is sticking to the young man' |
| c. | <u>Q</u> -menolela mokgalabje | 'He is overthrowing for the old man' |
| (4.3) a. | <u>Q</u> -tlo-lamolela mokgalabje | 'He will rescue for the old man' |
| b. | <u>Q</u> -tlo-mamarela lesogana | 'He will stick to the young man' |
| c. | <u>Q</u> -tlo-menolela mokgalabje | 'He will overthrow for the old man' |
| (4.4) a. | Re- <u>s</u> a-lamolela mokgalabje | 'We are still rescuing for the old man' |
| b. | Re- <u>s</u> a-mamarela lesogana | 'We are still sticking to the young man' |
| c. | Re- <u>s</u> a-menolela mokgalabje | 'We are still turning upside down for the old man' |
| (4.5) a. | Re- <u>l</u> e-lamolela lefsifsing (lephodisa) | 'We rescue for him in the dark (for the policeman)' |
| b. | Re- <u>l</u> e-mamarela lefsifsing (lesea) | 'We stick to it in the dark (to the baby)' |
| c. | Re- <u>l</u> e-menolela lefsifsing (lesoboro) | 'We overthrow for him in the dark (for the man)' |

