Nordic Journal of African Studies 26(4): 384–402 (2017)

The Status of Tone in Sesotho: A Production and Perception Study

Daan WISSING
and
Justus C. ROUX
North-West University, South Africa

ABSTRACT

Sesotho is generally described as a tonal language. This paper describes an investigation into the ability of a group of young speakers of Sesotho to perceive and produce tonal distinctions in Sesotho. In particular, it focuses on the status of the subject concord morpheme in the question phrases *o batlang*? (2SG; 'What do you want?') and *ó batlang*? (3SG/CL1; 'What does (s)he want?'). Thirty young mother tongue speakers aged between 15 and 17 participated in the study, which was conducted over a period of three months. The findings of both experiments clearly showed the functional use of tone by these speakers to be limited, or in some cases even totally absent. These results are suggestive of a system in a state of flux, perhaps indicating the start of an evolutionary process.

Keywords*: acoustic, intensity, evolution of tone, production, perception.*

1. INTRODUCTION

Sesotho[1] (sometimes referred to as Southern Sotho or simply Sotho) is mainly spoken in Lesotho and South Africa. Compared to other African tone languages the tonology of Southern Sotho is relatively well-studied. It is generally maintained that the Sotho languages (i.e. Sesotho sa Leboa, Setswana and Sesotho) are tone languages (Tucker 1929; Doke and Mofokeng 1957; Guma 1971; Roux 1983; Krüger and Snyman 1988; Poulos and Louwrens 1994). As in many Bantu languages, the tone system of the Sotho family underlyingly has a two-way contrast (L, H)[2] (Doke 1954; Cole 1955; Zerbian 2006).

Different aspects of tone that have been studied and reported on are tonal alternation (Khoali 1991; Mmusi 1992; Creissels 1996); high tone spreading (Creissels et al. 1997; Zerbian 2006), and the perception of tone (Mixdorff et al. 2011). Acoustic investigations of tone in Sotho include Khabanyane (1991), Selebeleng (1997), Zerbian and Barnard (2009), Mohasi et al. (2011) and

---

1 Sesotho is a Southern Bantu language, belonging to the Niger–Congo language family within the Sotho-Tswana branch of Zone S (S.30) https://en.wikipedia.org/wiki/Sotho_language.

2 L denotes Low Tone; H High Tone, as is customarily.

Wissing (2012), whilst Raborife (2012) presents a tone-labelling algorithm for Sotho.

As a tonal language, Sesotho uses pitch variations to convey lexical or grammatical meaning. Doke and Mofokeng (1957) provide the following Sesotho examples of minimal pairs in isolation: *ho báká* ('to repent') vs. *ho baka* ('to cause'); *ho tɛ́ná* ('to put on') vs. *ho tena* ('to disgust'), and in the case of grammatical meaning, minimal pair verb phrases such as: *o já dijɔ́* ('you eat food ') vs. *ɔ́ já dijɔ́* ('he eats food '). First-person singular subject markers can only be distinguished from similar third-person forms by means of tone in the sentences: *Ke motho* ('I am a person') vs. *Ké motho* ('It is a person'). Thus third-person forms generally carry high tone.

Wissing (2012) investigates the acoustic features of *ke motho* ('I am a person') vs. *ké motho* ('It is a person'), and to a lesser extent also *ho tɛ́ná* ('to put on') vs. *ho tena* ('to disgust') as produced by two groups of speakers, young and old, of both genders, inhabitants of the same region as those of the current study. In the case of *ke* in both structures the 2012 study reports limited support for the standard description of tone. The productions of adult speakers show clear support for the analysis of *ké* in the third-person form as high-toned. The same, however, is not true in the case of the production of the young speakers participating in that experiment. This finding is an indication that some kind of evolution could possibly be taking place in Sesotho. In the current study we concentrate on the language usage of young speakers of Sesotho with an expectation of corroborating the findings of Wissing (2012). We focus on another, similar grammatical construction as produced by a group of young speakers, and extend the study to their perceptions of the same structure. In this regard the vowels *o*- '2SG subject concord morpheme' (L) (in *wena o batlang*? 'What do you want?') and *ɔ́*- '3SG/CL1 subject concord morpheme' (H) (in *yena o batlang*? 'What does (s)he want?') as members of the minimal pair *o batlang*?/*ó batlang*? were involved in this investigation.[3] This pair assures optimal comparison of its members in both production and perception investigations. Furthermore the additional advantage exists, and will exploited here, that the interrogative form of the verb *batla* 'seek / want' + *(e)ng* ('what'), i.e. *batlang*? in the meaning of 'What do you want?' occurs with high frequency. Less frequent, though still possible, *Batlang*? in isolation could also serve as a one-word question in the sense of 'Want what?', which opens up the possibility of its serving as stimulus material for testing this short sentence in both intended meanings (*o*- 2SG and *ɔ́*- 3SG/CL1) in a perception task. This will be outlined in the subsequent sections.

The article is structured as follows: in section 2, different aspects of the experimental study are outlined. This comprises the construction of stimuli, the identification of a suitably representative speaker of Sesotho to make the

---

[3] Henceforth the abbreviations 2SG and 3SG/CL1 will be used only when referring to the complete descriptions of these two structures. In unambiguous cases only *o*- and *ó*- are used.

recordings to be used in the perception test, details about the recordings, the two experimental tasks, and also the processing of the information. The results of both Task 1 (the perception test) and of Task 2 (the production test) are provided in section 3, whilst a discussion of the results takes place in the fourth and final section. We conclude with some remarks on possible implications of the results for linguistic theory regarding the relationship between production and perception in general.

2. EXPERIMENTAL STUDY

2.1 GENERAL INFORMATION

A local intermediate school at Clarens, situated near Phuthaditjhaba in the deep rural South African Eastern Free State region, was chosen for the present study. Indications are that this region is very stable in terms of the inhabitants' migration patterns. This ensures optimal homogeneity of language usage by speakers of Sesotho.

2.1.1 Participants

Thirty listeners (eight boys and 22 girls)[4] took part in the perception task, and a few months later[5] fourteen of them (six boys and eight girls), who were available at the time, served as readers in the production task. One prototypical Sotho-speaking female (henceforth Speaker A), who had participated in previous projects and whose language use is regarded by peers as 'standard', served as the main reader of the stimulus materials. Three other adult speakers served as readers of the recordings used in the perception test. Their readings served as fillers in the perception task.

2.1.2 Experimental Design

The current study is confined to results of the production and perception tests of the grammatical constructions *o-batlang*?, *ó-batlang*? and of the isolated form *batlang*? Note that, parallel to these units, three structures with *motho* 'person'[6]

---

4 The uneven number of boys and girls is to be ascribed to the availability of participants at the time. Ideally, of course, the number of boys and girls should have been the same.

5 In order to avoid possible learning effects it was deemed wise not to run the production task too soon after the perception task. Practical considerations, such as obtaining an appointment with the school authorities, and travelling to Clarens, added to the time lapse between the two events.

6 These structures are *ke motho (I am a person), ké motho (it is a person),* and *motho* (person) (isolated).

as a basis were part of the stimulus materials, although they were not included in the analysis and results (see Appendix A).

The general design of the study comprises the following steps:

a. Construction of an appropriate set of stimuli, to be used in both tasks.
b. Identification of a suitably representative speaker of Sesotho (Speaker A) for recording of the stimuli in Task 1.
c. Recording by Speaker A and three other Sesotho speakers of readings of stimuli outlined in a' below.
d. Compilation of the collected stimuli for use in Task 1.
e. Running Task 1, the perception experiment.
f. Recording of readings of the stimulus structures (Task 2).
g. Processing of all information gathered.

In the following section these seven steps are explained in more detail.

a′: The grammatical structures are:

i. Wena o-batlang? (2SG)
ii. Yena ó-batlang? (3SG/CL1)
iii. X batlang? (2SG and also 3SG/CL1)

(i) and (ii) were read by Speaker A in two ways: once **with** inclusion of the pronouns *wena* or *yena*, and once **without** them.[7] X in (iii) represents either *wena o-* or *yena ó-*. Only one pair, viz. 3SG/CL1 + 2SG, was in fact included in the perception task.

b′: In a listening task performed by two other standard speakers of Sesotho to test the validity of Speaker A’s readings of a′ (i) and (ii), her productions were correctly classified 100% of the time. Results of the acoustic analysis of these productions (see Table 1 in the next section) confirmed their suitability as stimuli in the perception test described here. See g' (ii) below for the method used in extracting acoustic parameters.

c′ (& f′): The four speakers mentioned in the section *Participants* above were responsible for the recordings. In all recording tasks the same recording equipment and procedures were used and applied. A Shure unidirectional cardioid microphone was used in combination with a laptop. Care was taken to keep the distance between reader and microphone constant. Similar recordings made in previous studies, like the one reported by Wissing (2012), rendered results of adequate quality for the purposes of the present study too. The recordings were sampled at standard 44100Hz, 16-bit, stereo, in a Windows

[7] The reader was presented with the written forms *wena o batlang*? and *yena ó batlang*? In the first case, she had to keep *wena* in mind when reading only *o batlang*?; and similarly in the case of *yena ó batlang*?

Media Audio Bitrate format, 64 kbps, and then, via Adobe Audition, converted to normal wav format, also with sampling frequency of 44100 Hz, encoded in 16 bit linear PCM. In the second instance the same technical specifications were applicable, except that the recording was made directly in wav format. All recordings were highly suitable for the purposes of the present investigation.

d′: Eighteen sound files (in wav format) were prepared for this task (see Appendix A). A particular pair consists of either (1) a 2SG followed by 3SG/CL1, or in the reversed order, viz. (2) a 3SG/CL1 followed by 2SG, and (3) the isolated recording of *batlang*? Here only 3SG/CL1 followed by 2SG were implemented.

e′: The sound files were played via four loudspeakers evenly spread in the class room, within good hearing distance from every one of the 30 learners. One block of stimuli served as a trial run before commencement of the actual perception task.

In the perception task the 18 sound files were played pair by pair to the respondents. Listeners were instructed to identify the FIRST member of each pair that they heard. The two stimuli in each pair were separated by a one-second pause. Respondents had to indicate their choices (either WENA, YENA or NOT SURE) on the answering sheet (Appendix A). All participants reported their understanding of the task, as well as clear hearing of the stimuli.

The stimuli were played to the listeners in eight randomly ordered blocks, each block containing 18 sound pairs. Ample time was provided after every block. A stimulus pair was only played when all participants indicated their readiness to attend to the next one. If requested, a sound file was repeated. After each of the eight blocks a rest pause was allowed.

f′: Fourteen students who participated in Task 1 were available for reading the speech materials that had previously been employed. The same recording set-up and procedures for Task 2 (cf. c and f) were applied here. These recordings were made six weeks later in order to eliminate any learning effects from their participation in the perception task.

g′: The responses collected in Task 1 were scored, stored, and imported in a spreadsheet in the statistical package StatSoft® for statistical analysis.

In the case of Task 2 (the production test), the following procedures were followed, in compliance with generally acknowledged standards:

i. Oscillograms and spectrograms of the recordings were processed in Praat (Boersma and Weenink 2015). Speech signals were inspected auditorily and visually in order to determine and mark the vowel boundaries. Conventional segmentation criteria were followed (cf. Grabe and Low 2002, for example).

ii. ii. Vowels were annotated in Praat in such a way that Vowelyse (Van der Walt and Wissing 2003) could extract[8] and calculate the relevant acoustic information relevant for analysis, namely the first three vowel frequency formants (F1, F2, F3), and also the duration and intensity of the fundamental frequency (F0). Spurious formants and tone values are automatically detected and flagged by Vowelyse for later evaluation, and possible manual removal by the researcher. This is especially the case where Praat experiences problems in establishing a specific formant track, and then erroneously replaces the values with those of the next formant.

In the following section we report on the results of Task 1 and Task 2 separately.

3. RESULTS

3.1 TASK 1: THE PERCEPTION TEST

3.1.1 Acoustic Measurements

Here the acoustic results of analyses of the relevant vowels to be inspected in the perception test are provided. Note that due to the near-identical productions by Speaker A of *o*- and *ó*- in the structures with *wena* and *yena* respectively, and those without them (see (i) and (ii) in a' above), their acoustic measurements were merged, rendering a more unified and therefore manageable analysis and presentation of the results.

Table 1. *Acoustic results of the target vowels of the 2SG and 3SG/CL1 batlang? items, as produced by Speaker A, used in the perception task*. Measurements of F0, F1 – F3 are in Hertz (Hz); duration is in milliseconds (ms), and intensity is in dB. ba_2 and ba_3 represent the vowels of the first syllables of *batlang*.[9]

	o	*ó*	ba_2	ba_3
F1	409	468	751	768
F2	795	774	1630	1659
F3	2861	2994	2669	2842
F0	205	266	190	261
DUR	77	77	110	126
INTENS	82.2	83.0	84.1	85.2

8 Formant frequencies are average values of the mid 80% portion of each individual vowel.

9 Unlike the case of o- and ó-, the vowels of ba_2 and ba_3 should perhaps not be written as *a*- and *á*- respectively, given that these second of these vowels possibly acquires its H tone as a result of high tone spreading, which is a phonetic phenomenon. See Zerbian (2006) for the manifestation of high tone spread in Sesotho.

It is clear that pitch (F0) value of *ó-* is notably higher than that of *o-*; the Intensity of *ó-* also exceeds the measurement for *o-*. A discriminant function (DA) analysis shows a 100% correct classification in the case of these two parameters. A multiple regression analysis (MRA) result of $R^2 = 0.99$ is perfectly in line with that of the DA. Only for F1 and F2 the difference between ba_2 and ba_3 is 67%, and for Duration in the instance of *o-* and *ó-* only 50%.

Table 2 below provides acoustic details of the *batlang?* pair as read in isolation.

Table 2. *Acoustic results of the batlang? items (read in isolation), as produced by Speaker A, used in the perception task*. ba_2 and ba_3 represent the vowels of the first syllables of *batlang*.[10] F0, F1 – F3 values are in Hertz (Hz); duration is in milliseconds (ms.), and intensity is in dB.

	ba_2	ba_3
F1	749	773
F2	1629	1665
F3	2665	2848
F0	190	263
DUR	113	122
INTENS	82.3	85.3

The results of Table 1 and Table 2 are very similar, especially regarding F0 and intensity. This authenticates the perception results concerning these structures, as is reported in the next section.

3.1.2 The Perception Task

Here the results of the 30 participants are reported and interpreted. Table 3 contains the structures including the subject concord morpheme vowels ((i) and (ii) in a'), while Table 4's results are those of the reactions to stimuli where the subject concord morpheme vowels are excluded (that is, (iii) in a').

Table 3. *Reactions by 30 listeners to 16 o-batlang? structures as presented in 8 blocks.*

o batlang?		
2SG		
correct	incorrect	undecided
138	102	0

ó batlang?		
3 SG/CL1		
correct	incorrect	undecided
152	81	3

64.4% percent of the responses to the 3SG/CL1 stimuli were correct. This is somewhat of an improvement over the performance for the 2SG stimuli, where 57.5% of responses were correct. From these results alone it is not possible to

10 As is clear in the results, F0 (pitch, denoting tone) of ba_3 is markedly higher than that of ba_2. In this sense, the vowel may be phonetically transcribed as *á*. (Also see Footnote 9.)

tell whether only the tones of the subject concord morphemes in Speaker A's recordings to were responsible for the classifications, or whether other parameters (mentioned in Table 1), namely the formant frequencies and intensity, contributed to the listeners' decisions. Note, for example the finding reported there of the 100% correct classification of the vowels of the subject concord morphemes in terms of a discriminant analysis of Intensity (83% vs. 82.2% for 3SG/CL1 and 2SG respectively). An interesting thing about these findings is the fact that the group of young Sesotho-speaking listeners in question were not very convincingly able to identify either of the response items. This is in stark contrast to the 100% correct reactions by the two evaluators of Speaker A's recordings (mentioned in b' above). One should be wary not to equate the purpose and therefore the results of discrimination tests with those of identification tests, without the necessary caution, as in the present case.

The perception test involving *batlang?* in isolation produced interesting results. In this case only one pair of stimuli, namely 3SG/CL1 followed by 2SG was presented to the listeners. As before, they were requested to indicate what they judged the FIRST member to be. The task seemed to have been somewhat more complicated, as 24 responses were 'Not sure', that is 10% in total. On the other hand, 55.8% of the responses were correct. This is a clear indication that, even in the absence of the subject concord morpheme *ó*-, the acoustic information carried by the vowel of *ba* is sufficient for listeners to correctly identify this structure at a rate better than chance.

The different responses to the vowel of the first syllable of *batlang*? were markedly similar, here (85.2% vs. 84.1% for 3SG/CL1 and 2SG respectively). In perceptual terms the vowel in ba_3 is distinctly louder, and could most probably have assisted the listeners in their decision. The same applies to the much higher duration of ba_3 (126 ms) in comparison to only 110 ms. for ba_2. On the whole, this could mean that the vowels of the subject concord morphemes are not the sole carriers of second-person or third-person meaning. Rather, it seems probable that the entire structures, namely *o-batlang* and *ó-batlang*, were contributors to the degree of success in the identification task.

3.1.3 Responses of Individual Listeners

In instances such as these, where the results are rather moderate, it is deemed meaningful to inspect the responses per individual. Figure 1, constructed on the basis of the results given in Appendix B, reflects a rather diverse pattern.

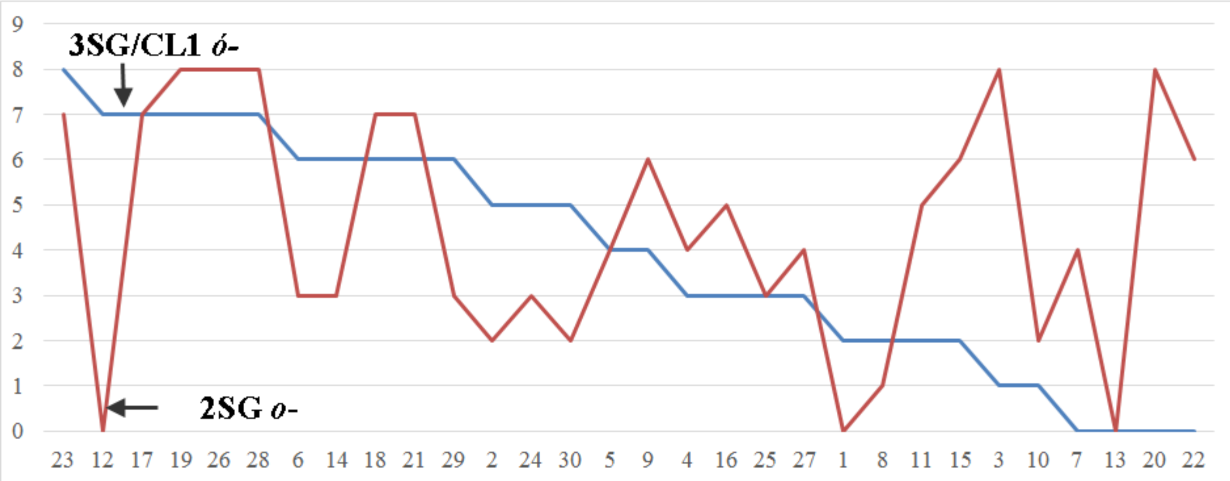

Figure 1. *Responses of the 30 participants (indicated on the x-axis) to the readings of Speaker A.* (The y-axis shows the number of correct responses. Responses regarding *ó-* are sorted from left to right, from maximum correct responses (8 = (100%) to minimum (0)).

The patterns for *o-* and *ó-* clearly show no similarity, but rather a notable difference. This is despite the fact that the average number of correct responses in both cases is around 4. Significant too is the number of respondents who were totally unable to perform this task. In the case of *o-*, three listeners failed to score a single correct response; four also responded similarly in the case of *ó-* (see x-axis, and Appendix B).

A general conclusion here is that it is not clear that tone is still functioning in the perceptual capacity of this group of young speakers. A question that should be addressed is whether there is a correlation between the participants' perception abilities and their production abilities. This issue, however, is beyond the aims of the present article.

3.2 TASK 2: THE PRODUCTION TEST

The basic production results obtained from fourteen of the thirty participants in the perception test reported above are provided in this section.[11] A precise comparison of the success per individual in the perception task with the individual's production ability should be most enlightening. An interesting hypothesis would be that readers' productions would be parallel to their degree of success in the perception task. Results could also possibly provide some interesting perspectives on the issue of whether good production necessitates good perception, or perhaps vice versa. However, the testing of such hypothesis lies beyond the scope of the present study.

---

11 Only these fourteen learners were available at the time of conducting of this experiment.

In Table 4 the results are provided of the acoustic analysis of the readings by fourteen readers of the structures described above (a′ i – ii). (Six readers' productions had to be discarded for a number of technical reasons, notably bad recordings.)

Table 4. *Results of four target vowels per acoustic parameters, as produced by 14 young Sesotho speakers*. (F0, F1 – F3 are measured in Hertz; Duration in seconds, and Intensity in dB. N = 329 – 330. Cohen's effect size values are given in the last two columns.[12])

							d-prime values	
		o-	*ó-*	(b)a_2	(b)a_3	All Grps	*o-* / *ó-*	a_2 / a_3
F0	Mean	206	210	185	94	199	0.07	0.16
	N	83	80	84	82	399		
	Std. Dev.	54	52	43	55	52		
Duration	Mean	0.078	0.086	0.114	0.117	0.099	0.38	0.21
	N	84	80	84	82	330		
	Std. Dev.	0.021	0.017	0.016	0.015	0.024		
F1	Mean	513	514	695	702	606	0.01	0.08
	N	84	80	84	82	330		
	Std. Dev.	75	71	90	81	122		
F2	Mean	953	921	1597	1630	1277	-0.12	0.16
	N	84	80	84	82	330		
	Std. Dev.	273	165	205	171	397		
F3	Mean	2976	2970	2807	2802	2889	-0.02	0.05
	N	84	80	84	81	329		
	Std. Dev.	350	320	348	308	341		
Intens	Mean	68	69	69	69	69	0.02	0.05
	N	84	80	84	82	330		
	Std. Dev.	7.1	6.7	5.8	5.9	6.4		

Except for Duration in the case of *o-* vs. *ó-* (76 ms resp. 86 ms), which translates into a d′ of 0.38 (a medium effect; translating to a moderate practical difference), all others clearly indicate what are practically highly non-significant differences (d′ = 0.01 – 0.21). F0, as an indicator of tone, and supposedly the main carrier of differences between 2SG and 3SG/CL1 structures in general, and

12 Effect size indexes in terms of d-values are calculated by the following formula (cf. Cohen, 1988): (Mean A minus Mean 2) / (largest standard deviation). This index is used for determining significance of the difference. d's smaller that 0.20 denotes no significant difference; *d* = 0.20 to 0.50 (equals a small effect); *d* = 0.50 to 0.08 (a moderate effect), and *d* > 0.80 (a large effect).

here, specifically of the differences between *o*- and *ó*-, does not contribute to any difference at all (d′ = 0.07). This finding is underscored by the results of an ANOVA for *o*- compared to *ó*- (F(7, 50) =.86, p=.54), and for ba_2 compared with ba_3 (F(7, 44)=.33, p=.94). Both differences are statistically highly insignificant.

When combining these parameters in one set, a discriminant function analysis renders a significantly dissimilar picture from that of Task 1. A strength scale of ba_3 > ba_2, and *o*- > *ó*- emerged, ranging from 65.3% of classifications being correct to only 46.1% being correct.

A more detailed analysis of the productions of *o*- and *ó*- per individual speaker in Task 2 generated the graph in Figure 2, specifically regarding tonal parameter F0. F0 is standardised to provide a direct comparison between the results of female and male speakers.

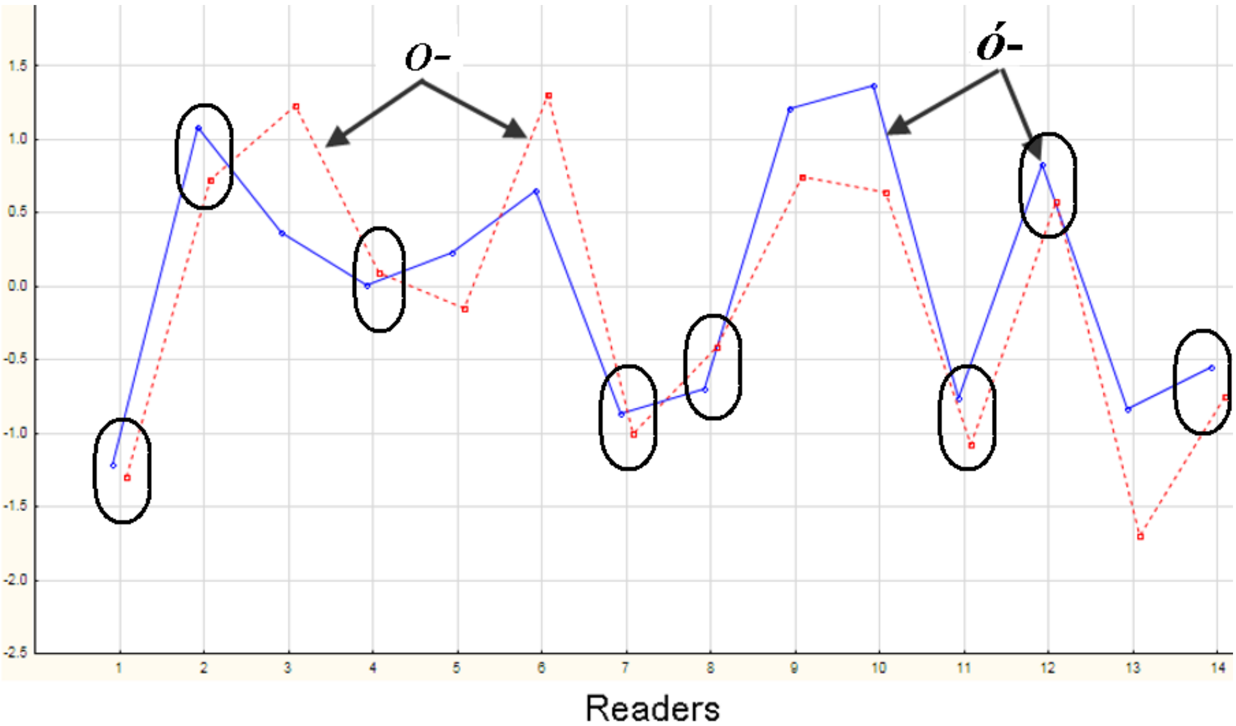

Figure 2. *Distribution and arrangement of standardised F0 of the two personal concords as produced by the fourteen young speakers of Sesotho*.

Firstly, in general, the diversity of the patterns per subject concord morpheme as well as per speaker is evident. Secondly, only three readers (9, 10, 13) produced *ó*- clearly as expected; the majority either did not differ noticeably from *o*- (enclosed in ellipses), or produced a tone higher than that of *ó*- (3, 6), instead of a lower one.

All in all, these findings once more reflect the very limited ability_of this group of speakers to master the phenomenon of tone in their language. In contrast to the MRA result of R^2= 0.99 reported for Speaker A's production of the same set of structures, R^2= 0.17 very clearly underlines the inability of the fourteen readers to correctly differentiate between them in their pronunciation.

3.2.1 Individual Participants' Productions and Perceptions

Where Sesotho speakers are perfectly proficient at both producing and perceiving grammatically significant tone distinctions, specifically in the *batlang*? constructions, one would expect an absolute correlation between pronunciation and interpretation. This applies even in the present investigation, where it was found that both production and perception turned out to be extremely arbitrary. Nevertheless it might be possible for some individual speakers – such as #9, #10 and #13 in Figure 2 – to correctly interpret clear productions as they did in the perception task.

Figure 3 depicts the perception performance of ten of the participants in the production task.[13]

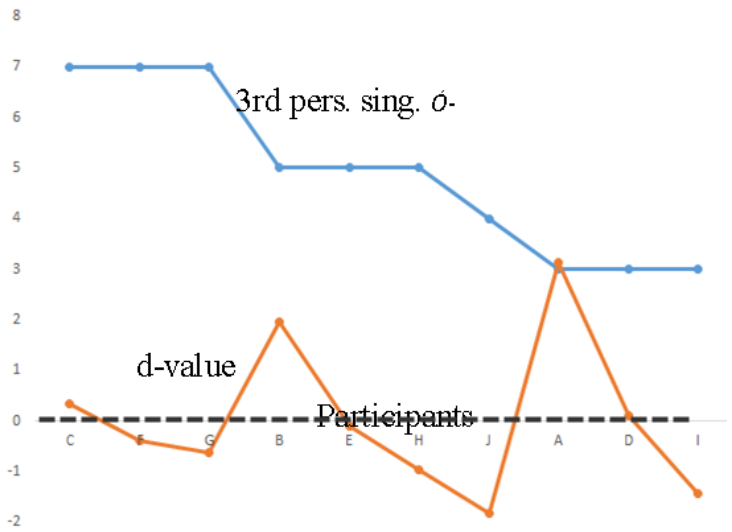

Figure 3. *Perception performance success of ten readers of Task 1. Maximum correct responses = 8. X-axis denotes d'-values*.

Note that five of the ten speakers (F, G, H, J, I – below the dashed line) erroneously interpreted *ó-* as *o-*. Furthermore, note the wide distribution of reactions, from 7 – 3. According to this graph only C's responses to the task might be viewed as clearly approaching what would be expected of a 'standard' speaker of Sesotho. These results therefore do not shed any light on the relation between production and perception of tone in the present investigation. Research into the ability of older persons might perhaps render quite a different picture, perhaps more in line or even completely in line with what could be expected, given current descriptions of the nature and role of tone in Sesotho.

---

[13] The productions of the other four participants in the production task were not available at the time of this *post hoc* analysis.

4. DISCUSSION AND CONCLUSION

The main aim of the present investigation was to determine whether tone as a conveyer of grammatical meaning in Sesotho is undergoing evolution, as the results of a previous study suggest. Its findings regarding the questions *o batlang*? ('What do you want?') and *ó batlang*? ('What does (s)he want?') underscore this possibility to an extent.

The overarching finding of our investigation with this particular group of mother-tongue speakers is the variation in perception and production of tone as a disambiguating factor in Sesotho. Although variation between individuals is expected of a sound system that is in flux, as was found in this study, and also in Wissing (2012), the factors that determine the production and perception patterns associated with an individual speaker remain unknown for now. More research has to be done on this.

Regarding the reactions to the sound files presented to the 30 participants, they show a preference for analysing the forms as 3SG/CL1. Strikingly even in the absence of the subject concord morphemes, that is, when listening *only* to *batlang*? in isolation, their correct reactions, albeit moderately positive, are nearly as successful as when *o*- or *ó*- is included. It is important to keep in mind the nature of auditory items offered to them in this task, namely readings by a competent speaker of Sotho: acoustically *o*- and *ó*- differ not only in tone but also markedly in intonation and duration. In the present research design it is not possible to ascertain the relative effect of these three parameters on their responses. This could be an interesting topic for future research.

An analysis of the success rate per individual listener in the perception task regarding the correct classification of 2SG or 3SG/CL1 structures shows no clear pattern. This agrees with the overall picture of a lack of complete mastery of this particular task.

Results for the productions of a subgroup of the 30 listeners that took part in the perception test corroborate, to a significant extent, the findings of Wissing (2012), whose results regarding the role of tone in short declarative sentences such as *Ke motho* ('I am a person') vs. *Ké motho* ('It is a person') were similar. Findings of the present investigation show that, in the production of the current group, it is only Duration, and not Tone, that differentiates (in favour of *ó*-) between the two sentence types to any significant extent .

Changes in tonal characteristics are not unique, as is evident from the results of the present study compared to standard descriptions of Sotho. For example, in a study on Contemporary Seoul Korean, Kang and Han (2013) found that the size of the F0 difference on vowels in some respects has increased over the course of the 20th century. Coetzee et al. (submitted) document an ongoing sound change in Afrikaans whereby an original plosive voicing contrast is being replaced by an innovative tonal contrast. On the other hand, tonal sound changes do not only imply an increase of tonal contrasts as in the two cases cited, or the establishment of such contrasts; tonal contrasts may very well also decrease, or even disappear.

Viewed in terms of the diverse models of sound change that have been proposed over time, the implications for the theory of language change of the present findings concerning the role of tone in Sesotho are all but clear. What is clear, though, is the present group of Sesotho speakers' rather limited degree of success if not actual incompetence both in producing and in perceiving tonal contrasts.

Literature differs regarding the precise link between perception and production. Kataoka (2011) provides a comprehensive overview of existing models of sound change that ranges from the primary importance of production as a change-initiating force to the primary importance of perception. In Paul's speaker-based theory (1886/1970) and others, auditory and articulatory feedbacks are identified as controlling speech production. On the other hand, Ohala (1981, 1989, 1993) in his Misperception Model, which is a listener-based model of change, ascribes the incongruity in pronunciation targets between a speaker and a listener to the listener's misperception. The Variation-Selection Model of Lindblom et al. (1995) attributes the same outcome to different modes of perception. Blevins's (2004) conception of speech perception portrays the first stage of change as a process where acoustic properties are mapped onto wrong phonetic representations owing to perceptual confusion. In this regard, Roux's (1983, 2003) proposed Perceptual Confusion Hypothesis may also be mentioned: this hypothesis suggests a potential confusion between vowel quality and tone in impressionistic phonetic descriptions of Sesotho. Finally, Beddor (2009) proposes a conceptual model of the role of listeners in sound change, in which variation in the perception grammar plays a central role. Kataoka's (2011) own investigation found no evidence for a perception-production link within subjects. The findings of the present study are in accordance with this conclusion. Although this kind of variation between individuals is expected of a sound system that is in flux, the factors that determine the production and perception patterns associated with an individual speaker remain unknown for now. Much more research into this intriguing and complicated topic still has to be done.

It may be possible to conclude that in the mouths and ears of these participants, tone is not alive and well in Sesotho. At the same time it is not possible or advisable to generalise about Sesotho as a language, but at least this study gives some reason to be cautious about generalising to the contrary, namely that Sesotho demonstrates a unitary tonal system, without variability or exception. The present findings might therefore be interpreted, with appropriate caution, as indicative of the presence of a change in the status of tone as one of the most prominent features of Sesotho.

The present findings could serve as a starting point for fruitful comparisons with results of similar investigations into the perception and production of older speakers. Such findings might shed light onto the possible loss, or partial loss, of tone as a robust feature of Sesotho.

ACKNOWLEDGEMENTS

This work is based on the research supported in part by the National Research Foundation (NRF) of South Africa (Grant Number 90529). We acknowledge that opinions, findings, conclusions and recommendations expressed are those of the authors and that the NRF accepts no liability whatsoever in this regard.

REFERENCES

Beddor, P. S. 2009.
A coarticulatory path to sound change. **Language** 85: 785–821.

Blevins, J. 2004.
Evolutionary phonology: The emergence of sound patterns. Cambridge: Cambridge University Press.

Boersma, P. and Weenink, D. 2015.
Praat: doing phonetics by computer. www.praat.org.

Cohen, J. 1988.
Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Earlbaum Associates.

Cole, D. T. 1955.
An introduction to Tswana grammar. London: Longmans.

Creissels, D. 1996.
Conjunctive and disjunctive verb forms in Setswana. **South African Journal of African Languages** 16(4): 109–115.

Creissels, D., Chebanne, A. M. and Nkhwa, H. W. 1997.
Tonal morphology of the Setswana verb. **LINCOM Studies in African Linguistics** 13: 21–44.

Doke, C. M. & Mofokeng, S. M. 1957.
Textbook of Southern Sotho. Cape Town: Longman.

Doke, C. M. 1954.
The Southern Bantu Languages. Cape Town: Oxford University Press.

Grabe, E. and Low, E. L. 2002.
Durational variability in speech and the rhythm class hypothesis. In: C. Gussenhoven and N. Warner (eds.), *Laboratory Phonology VII*, pp. 515–546. Berlin: Mouton de Gruyter.

Guma, S. M. 1971.
An outline structure of Southern Sotho. Pietermaritzburg: Shuter and Shooter.

Kataoka, Reiko. 2011.
Phonetic and cognitive bases of sound change. Ph.D. dissertation. University of California, Berkeley.

Khabanyane, K. E. 1991.
The five phonemic vowel heights of Southern Sotho: an acoustic and phonological analysis. In: *Working Papers of the Cornell Phonetics Laboratory*, No. 5, pp 1–36.

Khoali, B. T. 1991.
A Sesotho tonal grammar. Doctoral dissertation, University of Illinois at Urbana-Champaign.

Krüger, C. J. H. and Snyman, J. W. 1988.
The sound system of Setswana. Goodwood: Via Afrika.

Lindblom, B., Guion, S., Hura, S., Moon, S.-J. and Willerman, R. 1995.
Is sound change adaptive? **Rivista di Linguistica** 7: 5–37.

Mixdorff, H., Mohasi, L., Machobane, M. and Niesler, T. 2011.
A study on the perception of tone and intonation in Sesotho. In: *Proceedings of Interspeech 2011*, pp. 3181–3184.

Mmusi, S. O. 1992.
Obligatory contour principle effects and violations: the case of Setswana verbal tone. Doctoral dissertation, University of Illinois at Urbana-Champaign.

Mohasi, L., Mixdorff, H. and Niesler, T. 2011.
An acoustic analysis of tone in Sesotho. In: *Proceedings of International Conference of phonetic Sciences (ICPhS) XVII*, pp. 17–21.

Ohala, J. J. 1981.
The listener as a source of sound change. In: C. S. Masek, R. A. Hendrick, & M. F. Miller (eds.), *Papers from The 17th Regional Meeting of the Chicago Linguistics Society: Parasession on Language and Behavior*, pp. 178–203. Chicago: Chicago Linguistics Society.

1989 Sound change is drawn from a pool of synchronic variation. In: L. E. Breivik, & E. H. Jahr (eds.), *Trends in Linguistics, Studies and Monographs Language change: No. 43. Contributions to the study of its causes*, pp. 173–198. Berlin: Mouton de Gruyter.

1993 The phonetics of sound change. In: C. Jones (ed.), *Historical linguistics: problems and perspectives*, pp. 237–278. London: Longman.

Poulos, G. and Louwrens, L. J. 1994.
A Linguistic Analysis of Northern Sotho. Goodwood: Via Afrika.

Raborife, M. 2012.
Tone labelling algorithm for Sesotho. Doctoral dissertation, University of the Witwatersrand.

Roux, J. C. 1983.
Vokaalkwaliteit, toon en konsonantale invloede in Sotho. Unpublished research report. University of Stellenbosch, Stellenbosch.
2003 On the perception and description of tone in the Sotho and Nguni Languages. In: S. Kaji (ed.), *Cross-linguistic studies of tonal phenomena*, pp 155–176. Tokyo University of Foreign Studies.
Selebeleng, K. Z. 1997.
Phonetic and phonological aspects of vowel raising in SeSotho. Master's thesis, University of Stellenbosch.
Tucker, A. N. 1929.
The comparative phonetics of the Suto-Chuana group of Bantu languages. London: Longman, Green and Co.
Van der Walt, A. J. and Wissing, D. P. 2003.
Vowelyse: tools for acoustic analysis and optimising phonetic actions. http://www.puk.ac.za/fakulteite/lettere/HLT_Resources/.
Wissing, D. P. 2012.
On the many faces of tone in Southern Sotho: A case study. In: H. S. Ndinga-Koumba-Binza and S. E. Bosch (eds.), *Language Science and Language Technology in Africa*, pp 47-60. Stellenbosch: SUN Media.
Zerbian, S. 2006.
High tone spread in the Sotho verb. In: *Proceedings of the 35th Annual Conference on African Linguistics*, pp. 147–157.
Zerbian, S. and E. Barnard. 2009.
Realizations of a single high tone in Northern Sotho. **Southern African Language and Applied Language Studies** 27(4): 357–379.

About the authors: *Daan Wissing* is emeritus professor and senior researcher at the Centre for Text Technology, and the Research Unit: Language and Literature in the South African Context, North-West University, Potchefstroom, South Africa.

Justus Roux is emeritus professor and Director of the South African Centre for Digital Language Resources (SADiLaR), North-West University, Potchefstroom, South Africa.

APPENDIX A

Example of an answering sheet used in the perception task

EXERCISE	MARK ONE OF THREE		
1 o batlang	WENA	YENA	NOT SURE
2 ke motho	NNA	YENA	NOT SURE
3 o batlang	WENA	YENA	NOT SURE
4 ke motho	NNA	YENA	NOT SURE
5 o batlang	WENA	YENA	NOT SURE
6 ke motho	NNA	YENA	NOT SURE
7 o batlang	WENA	YENA	NOT SURE
8 o batlang	WENA	YENA	NOT SURE
9 ke motho	NNA	YENA	NOT SURE
10 o batlang	WENA	YENA	NOT SURE
11 ke motho	NNA	YENA	NOT SURE
12 ke motho	NNA	YENA	NOT SURE
13 o batlang	WENA	YENA	NOT SURE
14 o batlang	WENA	YENA	NOT SURE
15 ke motho	NNA	YENA	NOT SURE
16 batlang	WENA	YENA	NOT SURE
17 ke motho	NNA	YENA	NOT SURE
18 motho	NNA	YENA	NOT SURE

APPENDIX B

Responses of the 30 listeners on the stimulus material as read by Speaker A.

PARTICIPANT	3RD PERS CORRECT	2ND PERS CORRECT
1	2	0
3	1	8
4	3	4
5	4	4
6	6	3
7	0	4
8	2	1
9	4	6
10	1	2
11	2	5
12	7	0
13	0	0
14	6	3
15	2	6
16	3	5
17	7	7
18	6	7
19	7	8
2	5	2
20	0	8
21	6	7
22	0	6
23	8	7
24	5	3
25	3	3
26	7	8
27	3	4
28	7	8
29	6	3
30	5	2