

# Motivation for Corpus Selection

Mpho Raborife

Department of Applied Information Systems

University of Johannesburg

South Africa

March 3, 2025

## 1 High Tone Spread

1. (Ke (bala)⌘)C dinkgo  
1SG.SC count NP10.clay pot  
“I am counting the clay pots”
2. (Ke (labalabela)⌘)C tokoloho  
1SG.SC long NP9.freedom  
“I long for freedom”
3. (Re tla (páláma)⌘)C dithaba  
1SG.PL climb NP10.mountain  
“We climb mountains ”
4. (Re tla (hlómpha)⌘)C meétlo  
1PL.SC FUT respect NP4.tradition  
“We will respect traditions”
5. Kwena (é (fállá)⌘)C metsí  
NP9.crocodile SC9 emigrate NP4.water  
“The crocodile is migrating from its waters”
6. Leshodu (lé (tsúba)⌘)C kwaé  
NP5.thief SC5 smoke NP8.cigarettes  
AGR FUT AGR VPRS N NAGR N  
“He/she will buy you good food”
7. Moeti (ó tlá (tsamaya)⌘)C hoséng  
NP1.guest SC1 FUT leave ADV.morning  
“The guest will leave in the morning”
8. (Ke tla bá (rútá)⌘)C (ho (bíná)⌘)C difela  
1SG.SC FUT 3PL.OC teach INF sing NP8.hymns  
“I will teach them to sing hymns”

## 2 Minimal Pairs

1. ho tēna seaparo  
Ho tená seápáro (TR)  
“to put on clothes”

- ho tena motho  
Ho tená mothó (TR)  
“to annoy a person”
2. ho tshéla nóka  
Ho tshelá nóká (TR)  
“to cross a river”  
ho tshela metsi  
Ho tshela metsi (TR)  
“to pour water”
3. ho baka leratá  
Ho báká lerátá (TR)  
“to cause noise”  
ho báka sebe  
Ho báká sebe (TR)  
“to repent from sin”
4. ho lapá hoséng  
Ho lápá hoseng
5. “to be hungry in the morning”  
ho lapa mose  
Ho lápá mose (TR)  
“to patch a dress”
6. Yena ó mobe  
yená o mobe (TR)  
“He/she is ugly”  
Wena o mobe  
wená o móbe (TR)  
“You are ugly”
7. (Yena) ké ngwana wa háe  
yená ke ngwáná wá hác  
“He/she is his/her child”  
(Nná) ke ngwana wa háe  
nná ke ngwáná wá hác (TR)  
“I am his/her child”
8. Ké ngwana wa háe  
“He/she is his/her child”  
ké ngwáná wá hác (TR)
- Ké etetse haé  
“I am his/her child”

### 3 Transcriptions

#### 3.1 Minimal Pairs

There are 16 speakers, 8 from Gauteng townships (TS) and 8 from the Free State (FS). All speakers are under the age of 35 at the time of recording (2022). Last line is the final majority transcription. Below are the transcriptions

- Ho tshelá metsi (TS)  
Ho tshela metsi (FS)  
Final is a tie
- Ho tshelá nóká  
Ho tshelá nóká
- Ho lápa mose  
Ho lápa mose
- Ho lápa hoseng  
Ho lápa hoseng
- Ho tená seápáro  
Ho tená seápáro
- Ho tená mothó  
Ho tená mothó
- Ho báká sebe  
Ho báká sebe
- Ho báká lerátá  
Ho báká lerátá
- wená o móbe  
wená o móbe
- yená o mobe  
yená o mobe
- yená ke ngwáná wá hác  
yená ke ngwáná wá hác
- ké ngwáná wá hác  
ké ngwáná wá hác
- nná ke ngwáná wá hác  
nná ke ngwáná wá hác

## 4 Grammatical Tone

1. ((Pheháńg)⌘)C dijó  
cook.IMP NP8.food  
“Cook food”
2. Basotho ((nkínéléńg)⌘)C matsóho metsíńg  
NP2.sotho people dip.IMP NP6.hand NP4.water.LOC  
“The Sotho people, dip your hands in the water for me” literal meaning  
“The Sotho people, please forgive me” figurative meaning
3. (Ha ke (batlé)⌘)C batho  
NEG 1SG.SM want people  
“I do not want people”

4. (Ha ke (phehé)⌘)C dijó  
NEG 1SG.SM cook NP8.food  
“I am not cooking food”
5. (Ke (labálábétsé)⌘)C tokoloho  
1SG.SC yearn.PERF NP9.freedom  
I have yearned for freedom
6. (Ke ne ke (tuńmé)⌘)C motseng  
SC1 PST SC1 known.SUBJ NP3.community.LOC  
“I was well-known in the community”
7. ((Besá)⌘)C mollo, (ke (phehé)⌘)C dijó  
Make.VIMP NP3.fire 1SG.SC cook.SUBJ NP8.food  
“Make a fire so I can cook”
8. ((Thobá)⌘)C peló (o (tsamáyé)⌘)C morwetsana  
calm.VIMP OC1 go.SUBJ NP1.girl  
“Take heart and leave girl”

## 5 High Tone Spread Rule

In this study, we investigate the high tone spread rule amongst young speakers of Sesotho. This rule involves the spreading of an underlying high lexical tone to the adjacent syllable and is prevalent amongst Bantu languages [Demuth 1995]. We believe that by investigating its realisation amongst young speakers, we can contribute to the discourse of the evolution of tone in Sesotho. In Sesotho, this rule occurs within the clitic phrase and consists of the verb stem and all the clitics (prefixes) to the left of the verb stem [Khoali 1991].

In order to investigate the realisation of the high tone spread rule, we selected our corpus to illustrate instances where this rule is supposed to occur, according to the literature and instances where it should not occur. We focus on the clitic phrase, with polysyllabic verb stems. Monosyllabic verb stems show a complex tonal behaviour. For instance, the high tone of an object prefix does not spread to the adjacent monosyllabic verb stem as would be expected by the high tone spread rule [Khoali 1991]. It is for this reason that they have been excluded from the corpus. The sentences were selected to include the present principal tense as well as the future tense since they are relevant for the occurrence of the high tone spread rule.

1. Sentence 1 and 2 are used as a baseline of our study. No high tones are expected to be realised anywhere.
2. Sentence 3 and 4, the high tone spread rule is expected to occur with the verb stems.
3. Sentence 5 and 6 the rule is expected with the clitic phrase from the subject concord to the first syllable of a low-toned verb stem.
4. Sentence 7 a high tone is expected to occur with the clitic phrase, from the subject concord to the future tense marker.
5. Sentence 8 illustrates a more complicated sentence in our corpus, with a series of high tones adjacent to each other.

## 6 Minimal Pairs

We use minimal pairs to investigate lexical tone in heritage Sesotho speakers. We have selected frequently used words in our minimal pairs (words) in which the surrounding words provide context. This, we believe will assist the participants in disambiguating them by providing context. In Sesotho, singular subject markers can only be distinguished from similar third person forms by means of tone “ké motho - “It is a person” vs ke motho - “I am a person”. Wissing [2012] found that amongst the young speakers of Sesotho deviate from the standard description of tone in the production of singular subject markers vs. the third person form. In our study, we include this pair

and provide more context. This is also affirmed by Wissing and Roux [2017]. We have also included a similar construction in which the second person form can only be distinguished by tone from the 3rd person singular form; ó mobe - “he/she is ugly” vs. o mobe - “you are ugly”.

1. Phrases 1: 4 - verbs.
2. Phrase 5: distinguish 3SG from 2SG.
3. Phrase 6: distinguish 1SG from 2SG.
4. Phrase 7: distinguish possessive marker from noun.

## 7 Grammatical Tone Insertion Rule

In this study, we investigate the high tone spread rule amongst young speakers of Sesotho. This rule associates the second syllable of a verb stem with a high tone [Khoali 1991]. It occurs in the following context.

- The imperative mood.
- The negative verb form.
- The participial narrative past tense.
- The habitual verb form.
- The perfect tense.
- The subjunctive mood.

Once the grammatical tone has been realised in the second syllable in the above-mentioned contexts, it then spreads to the end of the phonological word (iterative high tone spread rule). Our corpus has sentences that include the imperative, the perfect tense as well as the subjunctive mood. In order to fully investigate the realisation of this rule as well as how its full consequence amongst our speaker, we use polysyllabic verb stems. The subjunctive mood is of particular interest in this case since it has been demonstrated in languages such as Spanish that heritage speakers demonstrate that the use of the subjunctive mood is reduced in cases where the indicative can be used [Sanchez Abchi and De Mier 2019; de la Fuente and Enriquez]. Our corpus consists of the imperative mood, negative verb form, perfect tense, and the subjunctive mood.

## References

- [de la Fuente and Enriquez ] A.A. de la Fuente and H. Enriquez, M.C. and Lacroix. Mood selection in relative clauses by french spanish bilinguals: Contrasts and similarities between l2 and heritage speakers. 3.
- [Demuth 1995] K. Demuth. Problems in the acquisition of tonal systems. In J. Archibald, editor, *The Acquisition of Non-linear Phonology*, pages 111–134. Lawrence Erlbaum Associates, Hillsdale, New York, 1995.
- [Khoali 1991] B. Khoali. *A Sesotho Tonal Grammar*. Theoretical Report, PhD Thesis, University of Illinois, 1991.
- [Sanchez Abchi and De Mier 2019] V Sanchez Abchi and V. De Mier. Verbs usage and narratives. an analysis of the verbal system in spanish heritage speakers productions. *Alfa, Sao Paolo*, 63(3):541–564, 2019.
- [Wissing and Roux 2017] D.P. Wissing and J.C. Roux. The status of tone in sesotho: A production and perception study. *Journal of African studies*, 26(4):541–564, 2017.
- [Wissing 2012] D. P. Wissing. On the many faces of tone in southern sotho: A case study. In H.S. Ndinga-Koumba-Binza and S.E. Bosch, editors, *Language Science and Language Technology in Africa*, pages 47–60. SUN Media, Stellenbosch, South Africa, 2012.