Word-level prosody in Sotho-Tswana

Sabine Zerbian[1]*,Etienne Barnard*[2]

[1] Department of Linguistics, University of Potsdam, Germany
[2] Human Language Technologies Research Group, Meraka Institute, CSIR, Pretoria, South Africa

szerbian@uni-potsdam.de, ebarnard@csir.co.za

Abstract

We introduce an algorithm to derive word-level tone assignments for the Sotho-Tswana languages. Prerequisite inputs are identified, and the steps to transform these inputs to tone assignments for each syllable are described. Manual implementation of the algorithm shows very good agreement with tone levels measured in a small Sotho-Tswana corpus.

Index Terms: Sotho-Tswana prosody, tone rules

1. Introduction

Southern Bantu languages are tone languages in which word-level pitch variations generally convey both lexical and grammatical meaning. In contrast to tone languages like Chinese, they are agglutinative languages, i.e. several morphemes are joined together in a word. Although most Southern Bantu languages only have two level tones, namely high tone (H) and low tone (L), modeling of their prosody is complicated by the agglutinative morphology, the significant influence of grammar and the occurrence of tone sandhi within and across words. Given the role of word-level prosody in processes such as semantic interpretation and the production of natural speech, it is important that a detailed and systematic account of the prosody be given. Such an account is complicated by the fact that tonal information is not indicated in the orthography of many Bantu languages (including those which are the focus of the current study).

Lexical tone is predictable neither from segmental form nor from morphological category (although the nominal domain is more tonal than verbs). Grammatical tone is largely predictable once a morphological analysis and labels are available. Previous work on intonation modeling in the South African Bantu language Zulu used statistical approaches (Louw, Davel & Barnard 2005; Kuun, Zimu, Barnard & Davel 2006; Govender, Barnard & Davel 2007; Levow 2009) without considering such factors in any detail.

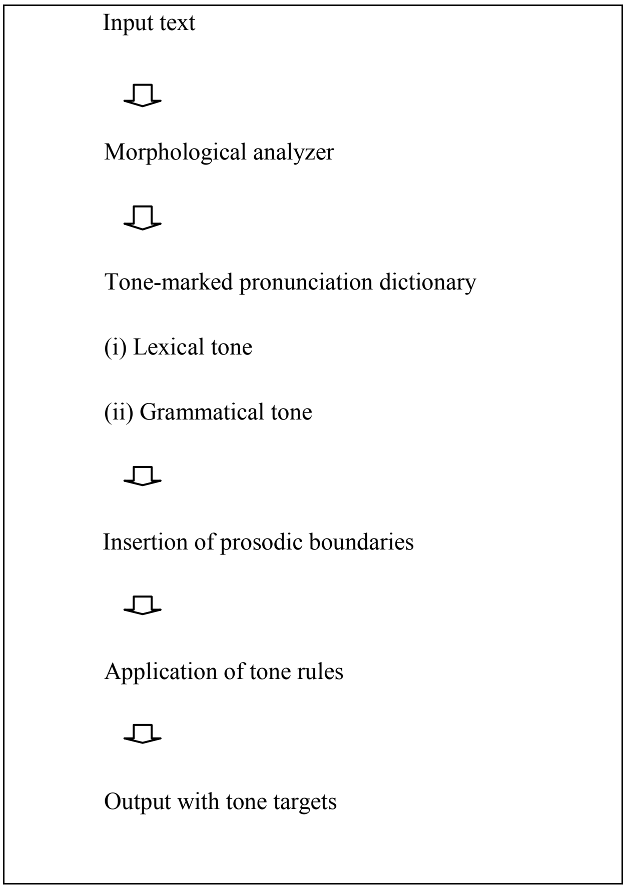
The current paper reports on a linguistically-informed approach to tone modeling in the Sotho-Tswana languages, which draws on tone-marked pronunciation dictionaries, morphological analysis, and tonal rules, which have been reported in the linguistics literature (e.g. Cole & Mokaila 1962; Chebanne *et al.* 1997) in order to generate labels for surface tone patterns using high (H) and low (L) tones, which eventually can be used to estimate pitch targets. The model has been pilot-tested on a small corpus of 15 hand-labelled sentences. The remainder of the paper uses the sentence in (1) from Northern Sotho as an example. Abbreviations used are the following: SG = singular; CL = noun class; NP = noun class prefix; PERF = perfect; SP = subject agreement prefix used in dependent clauses; ISP = subject agreement prefix used in independent clauses.

(1)	Ba	tloga	ba	thob-ile	kgobe	ka	mo-otlwa
	ISPCL2	leave	SPCL2	pierce-PERF	NP9.mealie	by	NP3-thorn

‘They are leaving after having pierced the boiled mealie with a thorn.’

Section 2 introduces the suggested modules for the prosody generation of Sotho-Tswana languages, schematized in (2). Section 3 reports on the evaluation of the model by means of a small, manually-labeled corpus. Section 4 concludes by placing this work into the larger context of sentence-level prosody.

(2) Step-wise prosody generation



Input text

Morphological analyzer

Tone-marked pronunciation dictionary

(i) Lexical tone

(ii) Grammatical tone

Insertion of prosodic boundaries

Application of tone rules

Output with tone targets

2. Prosody generation

2.1. Morphological analyzer / Shallow Parser

Tone in the Sotho-Tswana languages is dependent on morphology in several respects: first, whereas nouns can contrast tone on every syllable, verbs only contrast tone on their stem-initial syllable, which is either H or L. Second, each item in the closed set of grammatical morphemes has a fixed underlying tone. However, among these grammatical morphemes are many homographs which differ in tone only. A morphological analysis is necessary in order to assign the correct tone (cf. the subject concord *o* occurs as *o* (low tone) if it is IPPSP or as *ó* (high tone) if it is SP). Third, the tense, mood, and aspect (TMA) forms of verbs are associated with specific tone patterns (cf. *go-thobile* in (1) – 'to have pierced' occurs as *go-thobile* in the Independent Perfect Tense but as *go-thobílé* in the Dependent Perfect Tense). Again, only a morphological analysis of the verb word in its context can lead to the insertion of the correct tone pattern.

Thus, the verbs need to be labeled for categories such as dependent vs. independent tenses, present vs. past, positive vs. negative, etc. Also, subject concords need to be specified for these TMA-categories. Nominal constructions need to be broken down into noun class prefixes and stem.

(3) Morphological analyzer

ba	ISP CL2
tloga	Independent Present Tense
ba	SP CL2
thobile	Dependent Perfect Tense
kgobe	NPCL9.stem
ka	preposition
mo-otlwa	NPCL3-stem

2.2. Tone-marked pronunciation dictionary

The tone of vocabulary elements is largely unpredictable in tone languages. A tone-marked pronunciation dictionary is needed for the assignment of tone patterns for noun, verb and adjective stems. It is also needed for the assignment of the tone for grammatical morphemes such as subject concord, object concord, TMA-markers etc. As mentioned above, the correct assignment of tone for these morphemes, however, depends on the correct morphological analysis due to the many homographs in this category.

(4) Underlying tones (marked by underlining; from Ziervogel & Mokgokong 1975)

Ba̲ tloga ba̲ thobile kgobe ka̲ moo̲tlwa.

The tone-marked pronunciation dictionary should also provide the grammatical tone for the tenses, mood, and aspects (e.g. Cole & Mokaila 1962; Chebanne *et al.* 1997). Again, the correct assignment depends on the lexical tone patterns of stems in conjunction with the labels of TMA-categories.

There are three different (exceptional) tone patterns in Tswana. We only deal with the most frequent one here (the others occur in the Hortative and the Relative). This one occurs with imperatives, those negative infinitives using the negative particle *sa,* all negative verbal forms, verbal forms of the perfect (for exception see below), and habitual verbal forms. The pattern assigns a high tone on the second syllable of the verb stem (Khoali 1991).

(5) Grammatical tone (marked by italics)

Ba̲ tloga ba̲ thob*i*le kgobe ka̲ moo̲tlwa.]$_P$

2.3. Algorithm for prosodic phrase boundaries

Speech is not merely a process of concatenating stored phonological representations of lexical items: in production sounds change from their stored forms. These transformations may not always be influenced solely by the neighboring sounds, but may be due to the prosodic structure of an utterance which is influenced by syntactic considerations. Native speakers of a language are unaware of the underlying structures which determine the output form of speech but speech synthesizers will need to take them into account in order to sound natural.

At certain syntactic junctures so-called prosodic boundaries need to be inserted that influence the tonal realization of an utterance. The insertion of prosodic boundaries needs to precede the insertion of Grammatical tone from (5) because the presence or not of a boundary is crucial for the tone in the Perfect form. However, the prosodic boundaries only exert their influence on tone after all tone rules have applied.

A boundary is inserted before any punctuation mark[1] (comma or full stop), when a verb appears before certain adverbials, when a verb appears before locative prefixes, when a verb appears before a second high-toned subject concord in complex verb forms, and before a modifier (i.e. demonstrative, adjective, relative, possessive, enumerative). It is worth noting that these prosodic phrase (PP) boundaries are distinct from the intonation phrase (IP) boundaries. Whereas the IP boundaries determine the location of the penultimate lengthening and falling tone which are so salient in the Southern Bantu languages, the PP boundaries influence the action of tone rules, as discussed below.

(6) Prosodic boundaries

Ba̲ tloga ba̲ thob*i*le kgobe ka̲ moo̲tlwa]$_P$

[1] Though note that there exist no standardized punctuation rules for the Sotho-Tswana languages. Consequently, reliance on punctuation marks might be problematic for many texts.

2.4. Algorithm for tone rules

Bantu languages are well-known for their tone sandhi which changes underlying tone patterns. Three rules are addressed here, namely High Tone Spread, Iterative Grammatical Tone Spread and the Finality Rule.

The High Tone Spread Rule captures the observation that a high tone extends to the following syllable if this syllable is not itself followed by a high tone. This rule may not necessarily be a phonological rule but can be conceived as a phonetic implementation rule, as argued for in Zerbian & Barnard (2009).

Verb forms with a Grammatical High Tone (e.g. involving negation and certain tenses) on the second stem syllable show Iterative High Tone Spread onto all following syllables within the word (Khoali 1991; note that the lexical tone of the verb determines the tone of the first syllable in these cases.)

The Finality Rule captures the observation that (prosodic) phrase final syllables cannot be the target of High Tone Spread (Zerbian 2007). However, high tones that come from the lexicon are realized on final syllables.

(7) Tone rules (marked by accent)

(7a) Specific rule of Iterative High Tone Spread

Ba̲ tloga ba̲ thob*i*lé kgobe ka̲ moo̲tlwa.]$_{\text{P}}$

(7b) General rule of High Tone Spread

Ba̲ tlóga ba̲ thob*i*lé kgobe ka̲ moo̲tlwá.]$_{\text{P}}$

(7c) Finality Rule

Ba̲ tlóga ba̲ thob*i*lé kgobe ka̲ moo̲tlwa.]$_{\text{P}}$

The resulting tone pattern can then be translated into pitch targets, as in (8).

(8) Resulting surface tone (high tones marked by acute)

Bá tlóga bá thobílé kgobe ká moótlwa.

H H L H L H H L L H LH L

3. Evaluation of the model

A small test corpus was developed for the evaluation of this prosody generation model. It comprises 15 sentences taken from the Northern Sotho TTS corpus of the HLT Research Group of the Meraka Institute at the CSIR (Van Niekerk and Barnard, 2009). The sentences were selected according to the following criteria: They must not contain proper nouns and/or loan words as the tone patterns for these words are not available. Verb forms showing the Potential Mood (using *-ka-*) have been excluded as the tonology of these forms is subject to considerable dialectal variation across the Sotho-Tswana languages (cf. Chebanne *et al*. 1997, Cole & Mokaila 1962, Lombard 1976).

All syllables of the 15 sentences of the corpus were labeled for tone by three labelers independently of each other. The labelers are sensitive to issues of tone but differ in their background and experience regarding Bantu tone. The individual labels were based on perception of the recorded sentences, acoustic analysis using the *Praat* software package (Boersma 2001) or both (cf. Govender et al. 2007 for a different labeling procedure). The labeled sentences were compared across all three labelers, which revealed unanimous agreement on the tone labels in 72.3% of the cases (196 out of 271 syllables). A final transcription was generated which is informed by the majority vote, i.e. it is based on the tone label that at least two labelers agreed on. It needs to be noted that the corpus was compiled for purposes of speech synthesis so the speaker was asked to produce a rather monotonous speech melody. This flat intonation might have been one of the reasons for cases of disagreement between labelers' decisions.

The transcribed tone patterns for these 15 sentences were then compared to the predicted tone labels based on the step-wise prosody generation presented in the current paper. Manual morphological analysis, parsing and boundary assignment were performed, and the underlying tone values of all morphemes and stems were obtained from a standard dictionary (Ziervogel & Mokgokong 1975). Of the 271 labels, 257 matched, resulting in 94.8 % accuracy.

4. Conclusion

We have provided an algorithmic approach to the prediction of tone labels in Sotho-Tswana, which is able to predict observed tone levels with high accuracy in a small test corpus. Given the very good match obtained, work on an automated algorithm for the implementation of tone rules has started (Raborife 2009).

The paper has concentrated mainly on word-level prosody with some tonal alterations occurring at sentence-level to indicate phrasing (cf. Finality rule). However, a language's overall intonation includes pitch alternations at the utterance-level for the indication of sentence type and emphasis. From the existing literature it seems that there is little tonal sentence-level intonation in Sotho-Tswana other than declination (language universal, expected to occur over the course of an utterance) and restricted question prosody. Jones *et al.* (2001a, b) report on a specific question prosody in yes/no-questions in Xhosa (Nguni) which includes a raised overall register for these questions and the absence of penultimate lengthening. Impressionistic observations suggest this pattern for Sotho-Tswana as well. Crucially, focus or discourse-new information is not marked prosodically in Sotho-Tswana (Zerbian 2007) in contrast to English. However, non-tonal cues are used for intonational purposes (cf. Hyman & Monaka 2008). Thus, the Sotho-Tswana languages seem to have what Michaud (2006) calls 'calculated prosody', which occurs in languages in which tone serves complex morphophonological functions and whose prosodic structure relies on the calculation of tone sequences in a categorical way. These languages contrast with languages like English or Chinese which have fewer elements of categorical tonal calculation and in which intonation appears in a largely non-categorical way.

As mentioned in the Introduction, the work described here is of both linguistic and technical importance. We conclude by discussing some of its implications for speech technology in the Sotho-Tswana languages. For automatic speech recognition (ASR), accurate tone modeling is not likely to play a significant role in the near future. On the one hand, tonal information is not indicated in the orthography, so that transcription and dictation systems are not required to capture such information. On the other hand, the distinctions that are expressed by tonal differences in otherwise indistinguishable utterances are unlikely to be important in Spoken Dialog Systems, which rely on the availability of substantial contextual information in any case. (That is, sufficiently sophisticated Spoken Dialog Systems, which may be misled by misinterpretation of tonal information, are still a long way off.) For speech synthesis, or text-to-speech (TTS) systems, the situation is quite different. TTS systems cannot but generate *some* pitch contour, and the naturalness of a TTS system depends strongly on the properties of this contour (along with other prosodic factors such as pauses, rhythm, etc.) Hence, detailed and accurate prosodic models are a prerequisite for natural-sounding TTS, and such systems are likely to be the most important application domain of this work in years to come.

In order for our work to be useful for TTS systems, two additional components must be developed. At the front end, morphological analysis and parsing must be automated, and at the back end, fundamental-frequency (F0) contours must be computed from the tone assignments made by our system. Much progress has been made on the former challenge in recent years (see, for example, Faasz 2009, Pretorius et al. 2009), and the latter challenge is being addressed in our ongoing research. Electronic pronunciation dictionaries for the Sotho-Tswana languages are available (Davel 2009), and the incorporation of tone information into these dictionaries is also a current project.

5. References

Boersma, P. 2001. *Praat, a system for doing phonetics by computer*. Amsterdam: Glott International.

Chebanne, A.M., Creissels, D. & Nkhwa, H.W. 1997. *Tonal Morphology of the Setswana Verb*. Munich: LINCOM Europe.

Cole, D.T. & Mokaila, D.M. 1962. *A Course in Tswana*. Washington: Georgetown University Press.

Davel M. and Martirosian O.M. 2009. Pronunciation dictionary development in resource-scarce environments. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*, Brighton UK: 2851-2854.

Faasz, G., Heid, U., Taljard, E. & Prinsloo, D.J. 2009 Part-of-Speech tagging of Northern Sotho: Disambiguating polysemous function words. *Proceedings: EACL 2009 Workshop on Language Technologies for African Languages (AfLaT 2009)* Athens, Greece.

Govender, N., Barnard, E. & M. Davel. 2007. Pitch Modelling for the Nguni Languages. SACJ 38: 28-39.

Hyman, L. M. & K. C. Monaka. 2008. Tonal and Non-Tonal Intonation in Shekgalagari. UC Berkeley Lab Annual Report: 269-288.

Jones J., Louw, J.A. & Roux, J.C. 2001a. Perceptual experiments on Queclaratives in Xhosa. S*AJAL*, supplement 36: 19–31.

Jones J., Louw, J.A. & Roux, J.C. 2001b. Queclaratives in Xhosa: an acoustic analysis. *SAJAL*, supplement 36: 3–18.

Khoali, B.T. 1991. A Sesotho Tonal Grammar. Unpublished PhD thesis, University of Illinois at Urbana-Champaign.

Kuun, C., Zimu, V., Barnard, E. & M. Davel. 2006. Statistical investigations into isiZulu intonation. ISCA Tutorial and Research Workshop, Stellenbosch, South Africa.

Levow, G.-A. 2009. Assessing Context and Learning for isiZulu Tone Recognition. *Proceedings of Interspeech, Brighton, UK*.

Lombard, D.P. 1976. Aspekte van Toon in Noord-Sotho. Unpublished PhD thesis, University of South Africa.

Louw, J.A., Davel, M. & E. Barnard. 2005. A general-purpose IsiZulu speech synthesizer. South African Journal of African Languages 25: 92-100.

Michaud, A. 2006. Replicating in Naxi (Tibeto-Burman) an Experiment Designed for Yorùbá: An Approach to 'Prominence-Sensitive Prosody' vs. 'Calculated Prosody'. In *Proceedings of Speech Prosody 2006*, Dresden: 819-822.

Pretorius, R., Berg, A., Pretorius, L. & B. Viljoen 2009. Setswana Tokenisation and Computational Verb Morphology: Facing the challenge of a disjunctive orthography. Proceedings of the First Workshop on Language Technology for African Languages (AfLaT 2009), Association for Computational Linguistics, Athens, Greece, p. 66-73. Available at http://www.aclweb.org/anthology/w09-0710.

Raborife, M.I. 2009. The Implementation of Sesotho Tonal Rules in a Text-to-Speech System. Honours Report, School of Computer Science, University of the Witwatersrand, Johannesburg.

Van Niekerk, D.R. & Barnard, E. 2009 Phonetic alignment for speech synthesis in under-resourced languages In *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009),* Brighton, UK: 880-883

Zerbian, S. 2007. Investigating prosodic focus marking in Northern Sotho. In Hartmann, K., Aboh, E. & Zimmermann, M. (eds.) *Focus strategies: evidence from African languages*. Berlin: Mouton de Gruyter, pp. 55-79.

Zerbian, S. & E. Barnard. 2009. Realizations of a single high tone in Northern Sotho. SALALS. 27(4): 357–379

Ziervogel, D. & P. C. Mokgokong. 1975. *Groot Noord-Sotho woordeboek*. Pretoria: Van Schaik.